Volume Licensing

Service Level Agreement for Microsoft Online Services

November 1, 2025

Microsoft

Table of Contents

TABLE OF CONTENTS....2
INTRODUCTION....4
GENERAL TERMS....5
SERVICE SPECIFIC TERMS....7
MICROSOFT DYNAMICS 365....7
DYNAMICS 365 BUSINESS CENTRAL....7
DYNAMICS 365 COMMERCE....7
DYNAMICS 365 CONTACT CENTER....8
DYNAMICS 365 CUSTOMER INSIGHTS....8
DYNAMICS 365 CUSTOMER SERVICE ENTERPRISE; DYNAMICS 365 CUSTOMER SERVICE PROFESSIONAL; DYNAMICS 365 CUSTOMER SERVICE INSIGHTS; DYNAMICS 365 FIELD SERVICE; DYNAMICS 365 MARKETING 8
DYNAMICS 365 GUIDES....9
DYNAMICS 365 HUMAN RESOURCES....9
DYNAMICS 365 INTELLIGENT ORDER MANAGEMENT....9
DYNAMICS 365 REMOTE ASSIST....10
DYNAMICS 365 SALES ENTERPRISE; DYNAMICS 365 SALES PROFESSIONAL 10
DYNAMICS 365 SUPPLY CHAIN MANAGEMENT; DYNAMICS 365 FINANCE; DYNAMICS 365 PROJECT OPERATIONS 11
OFFICE 365 SERVICES....11
DUET ENTERPRISE ONLINE....11
EXCHANGE ONLINE....12
EXCHANGE ONLINE ARCHIVING....14
EXCHANGE ONLINE PROTECTION....14
MICROSOFT MYANALYTICS....15
MICROSOFT STREAM (CLASSIC)....15
MICROSOFT TEAMS....15
MICROSOFT 365 APPS FOR BUSINESS....16
MICROSOFT 365 APPS FOR ENTERPRISE....16
OFFICE 365 ADVANCED COMPLIANCE....16
OFFICE ONLINE....17
OFFICE 365 VIDEO....17
ONEDRIVE FOR WORK OR SCHOOL....17
PROJECT....17
SHAREPOINT ONLINE....18
MICROSOFT TEAMS – CALLING PLANS, TEAMS PHONE AND AUDIO CONFERENCING 18
VIVA ENGAGE....19
MICROSOFT TEAMS – VOICE QUALITY....19
WORKPLACE ANALYTICS....19
MICROSOFT AZURE SERVICES AND PLANS....20
MICROSOFT ENTRA ID....20
AZURE ACTIVE DIRECTORY B2C....20
MICROSOFT ENTRA DOMAIN SERVICES....20
ANALYSIS SERVICES....21
AZURE API FOR FHIR....21
API CENTER SERVICES....21
API MANAGEMENT SERVICES....22
APP CENTER....22
APP CONFIGURATION....23
APP SERVICE....24
APPLICATION GATEWAY....25
APPLICATION GATEWAY FOR CONTAINERS....25
APPLICATION INSIGHTS....26
AZURE ARC....26
AUTOMATION....26
AZURE BACKUP....27
AZURE BASTION....28

Batch....28
BizTalk Services....28
Azure Bot Service....29
Azure Cache for Redis....29
Azure Managed Redis....30
Azure Chaos Studio....31
Azure Cloud HSM....31
Cloud Services....32
Azure AI Search....32
Azure AI Services....33
Azure Communication Gateway....33
Azure Communication Services....34
Azure Confidential Ledger....34
Azure Container Apps....35
Azure Container Instances....35
Azure Container Registry....35
Content Delivery Network (CDN)....36
Azure Cosmos DB....36
Data Catalog....40
Azure Data Explorer (Kusto)....40
Azure Data Factory....41
Data Lake Analytics....41
Data Lake Storage Gen1....42
Azure Database for MariaDB....42
Azure Database for MySQL....42
Azure Database for PostgreSQL....43
Azure Databricks....45
Microsoft Azure Data Manager for Energy....45
Azure DDoS Protection....45
Azure Defender....46
Defender External Attack Surface Management....46
Azure Dev Ops....47
Microsoft Dev Box....47
Azure Digital Twins....48
Azure DNS....48
Azure DNS Private Resolver....49
Azure DocumentDB (with MongoDB compatibility)..49
Elastic San SLA....50
Event Grid....50
Geo Redundant Storage (GRS) - Priority Replication51
Event Hubs....51
Azure ExpressRoute....52
Azure ExpressRoute Traffic Collector....53
Azure Files Premium Tier....53
Azure Firewall....54
Azure Fluid Relay....54
Azure Front Door and Azure Front Door (classic) 55
Azure Functions....55
Global Secure Access....56
HDInsight....56
Azure Health Data Services (excluding the MedTech Service) 57
Health Bot....57
Azure Information Protection....58
Azure IoT Central....58
Azure IoT Hub....58
Key Vault....59
Azure Key Vault Managed HSM....59
Azure Kubernetes Service (AKS)....60
Azure Lab Services....61
Azure Load Balancer....61
Azure Load Testing....62
Log Analytics (Query Availability SLA)....62

Logic Apps....62
Azure Machine Learning....63
Azure Managed Grafana....63
Azure Managed Instance for Apache Cassandra....64
Azure Maps....65
Media Services....65
MedTech service....67
Microsoft Fabric....68
Microsoft Genomics....68
Microsoft Sentinel....69
Mobile Services....69
Azure Monitor....69
Azure NetApp Files....70
Network Watcher....71
Notification Hubs....71
On Demand Capacity Reservations for Azure Virtual Machines72
Azure AI Foundry Models....72
Object Replication - Priority Replication....73
Azure Operator Service Manager....74
Azure Orbital Ground Station....75
Microsoft Playwright Testing....75
Azure Private Link....76
Microsoft Purview....76
Azure Red Hat OpenShift....76
Remote Rendering....77
Azure Route Server....77
SAP HANA on Azure Large Instances....78
Scheduler....79
Service-Bus....79
Azure SignalR Service....80
Azure Site Recovery....81
Spatial Anchors....82
Azure Spring Apps....82
Azure SQL Database....82
Azure SQL Managed Instance....84
SQL Server Stretch Database....85
Static Web Apps....85
Storage Accounts....86
StorSimple....88
Azure Stream Analytics....88
Azure Synapse Analytics....89
Azure Time Series Insights....90
Traffic Manager Service....90
Trusted Signing....91
Virtual Machines....91
Azure Virtual Network Manager....93
Azure Virtual WAN....93
Azure VMware Solution....93
Azure VMware Solution by CloudSimple....94
Azure VNet NAT....95
Virtual Network Gateway....95
Azure Web PubSub....96
Windows 10 IoT Core Services....96
Other Online Services....97
Microsoft Defender for Identity....97
Microsoft Defender for IoT....97
Bing Maps Enterprise Platform....97
Bing Maps Mobile Asset Management....98
Microsoft Cloud App Security....98
Dragon Copilot....98
Microsoft Power Automate....99
Microsoft Power Pages....99

Microsoft Intune.....100
Microsoft Kaizala Pro.....100
Microsoft Power Apps.....100
Microsoft Copilot Studio.....101
Microsoft Sustainability Manager.....101
Minecraft: Education Edition.....101
Power BI Embedded.....102
Power BI Premium.....102
Power BI Pro.....102
Azure AI Translator.....103
Microsoft Defender for Endpoint.....103
Universal Print.....103
Windows 365.....104

APPENDIX A – SERVICE LEVEL COMMITMENT FOR VIRUS DETECTION AND BLOCKING, SPAM EFFECTIVENESS, OR FALSE POSITIVE.....105

Introduction

About this Document

This Service Level Agreement for Microsoft Online Services (this "SLA") is a part of your Microsoft volume licensing agreement (the "Agreement"). Capitalized terms used but not defined in this SLA will have the meaning assigned to them in the Agreement. This SLA applies to the Microsoft Online Services listed herein (a "Service" or the "Services"), but does not apply to separately branded services made available with or connected to the Services or to any on-premise software that is part of any Service.

If we do not achieve and maintain the Service Levels for each Service as described in this SLA, then you may be eligible for a credit towards a portion of your monthly service fees. We will not modify the terms of your SLA during the initial term of your subscription; however, if you renew your subscription, the version of this SLA that is current at the time of renewal will apply throughout your renewal term. We will provide at least 90 days' notice for adverse material changes to this SLA. You can review the most current version of this SLA at any time by visiting https://aka.ms/CSLA.

Previews and Online Services and/or service tiers provided free of charge are not included or eligible for SLA claims or credits.

Prior Versions of this Document

This SLA provides information on Services currently available. Earlier versions of this document are available at http://www.microsoftvolumelicensing.com. To find the needed version, a customer may contact its reseller or Microsoft Account Manager.

Clarifications and Summary of Changes to this Document

Below are recent additions, deletions and other changes to this SLA. Also listed below, are clarifications of Microsoft policy in response to common customer questions.

Additions/Updates	Deletions
Azure Cosmos DB	Azure Private 5G Core
Geo Redundant Storage (GRS) – Priority Replication	
Object Replication – Priority Replication	

Table of Contents / Definitions

General Terms

Definitions

"**Applicable Period**" for metered Pay As You Go Services (e.g. Azure Virtual Machines) means the 30 days prior to and including the first day of the Incident for which a Service Credit is owed. For other services, (e.g. M365 E3) "Applicable Period" means the calendar month in which a Service Credit is owed.

"**Applicable Service Fees**" means the total fees actually paid by you for a Service that are applied to the Applicable Period in which a Service Credit is owed.

"**Downtime**" is defined for each Service in the Services Specific Terms below. Downtime does not include Scheduled Downtime. Downtime does not include unavailability of a Service due to limitations described below and in the Services Specific Terms.

"**Error Code**" means an indication that an operation has failed, such as an HTTP status code in the 5xx range.

"**External Connectivity**" is bi-directional network traffic over supported protocols such as HTTP and HTTPS that can be sent and received from a public IP address.

"**Incident**" means (i) any single event, or (ii) any set of events, that result in Downtime.

"**Management Portal**" means the web interface, provided by Microsoft, through which customers may manage the Service.

"**Scheduled Downtime**" means periods of Downtime related to network, hardware, or Service maintenance or upgrades. We will publish notice or notify you at least five (5) days prior to the commencement of such Downtime.

"**Service Credit**" is the percentage of the Applicable Service Fees credited to you following Microsoft's claim approval.

"**Service Level**" means the performance metric(s) set forth in this SLA that Microsoft agrees to meet in the delivery of the Services.

"**Service Resource**" means an individual resource available for use within a Service.

"**Success Code**" means an indication that an operation has succeeded, such as an HTTP status code in the 2xx range.

"**Support Window**" refers to the period of time during which a Service feature or compatibility with a separate product or service is supported.

"**User Minutes**" means the total number of minutes in an Applicable Period, less all Scheduled Downtime, multiplied by the total number of users.

Terms

Claims

In order for Microsoft to consider a claim, you must submit the claim to customer support at Microsoft Corporation including all information necessary for Microsoft to validate the claim, including but not limited to: (i) a detailed description of the Incident; (ii) information regarding the time and duration of downtime; (iii) affected resource names; (iv) the number and location(s) of affected users; and v) a description of errors that occurred during the incident. Failure to provide the required information will result in claims being rejected. For detailed guidance on monitoring and logging Azure resources, please refer to the Azure Monitor log analytics queries tables" article or any successor articles on Microsoft Learn. This resource provides essential data and insights needed for Azure claims.

For a claim related to Microsoft Azure, we must receive the claim within 60 days of the Incident. For claims related to all other Services, we must receive the claim by the end of the Applicable Period following the month in which the Incident occurred. For example, if the Incident occurred on February 15th, we must receive the claim and all required information by March 31st.

We will evaluate all information reasonably available to us and make a good faith determination of whether a Service Credit is owed. We will use commercially reasonable efforts to process claims promptly following an investigation, typically within forty-five (45) days of receipt of the claim. You must be in compliance with the Agreement in order to be eligible for a Service Credit. If we determine that a Service Credit is owed to you, we will apply the Service Credit to your Applicable Service Fees.

If you purchased more than one Service (not as a suite), then you may submit claims pursuant to the process described above as if each Service were covered by an individual SLA. For example, if you purchased both Exchange Online and SharePoint Online (not as part of a suite), and during the term of the subscription an Incident caused Downtime for both Services, then you could be eligible for two separate Service Credits (one for each Service), by

submitting two claims under this SLA. In the event that more than one Service Level for a particular Service is not met because of the same Incident, you must choose only one Service Level under which to make a claim based on the Incident. Unless as otherwise provided in a specific SLA, only one Service Credit is permitted per Service for an Applicable Period.

Service Credits

Service Credits are your sole and exclusive remedy for any performance or availability issues for any Service under the Agreement and this SLA. You may not unilaterally offset your Applicable Service Fees for any performance or availability issues.

Service Credits apply only to fees paid for the particular Service, Service Resource, or Service tier for which a Service Level has not been met. In cases where Service Levels apply to individual Service Resources or to separate Service tiers, Service Credits apply only to fees paid for the affected Service Resource or Service tier, as applicable, and only with respect to the actual downtime and not the overall outage duration. The Service Credits awarded in any billing month for a particular Service or Service Resource will not, under any circumstance, exceed your monthly service fees for that Service or Service Resource, as applicable, in the Applicable Period. Service Credits will not be awarded to compensate for any other forms of loss, including but not limited to lost revenue, operational costs, or any indirect losses experienced by you or the end-users.

Limitations

This SLA and any applicable Service Levels do not apply to any performance or availability issues:

1. Due to factors outside our reasonable control (for example, natural disaster, war, acts of terrorism, riots, government action, or a network or device failure external to our data centers, including at your site or between your site and our data center);
2. That result from the use of services, hardware, or software not provided by us, including, but not limited to, issues resulting from inadequate bandwidth or related to third-party software or services;
3. That results from failures in a single Microsoft Datacenter location, when your network connectivity is explicitly dependent on that location in a non-geo-resilient manner;
4. Caused by your use of a Service after we advised you to modify your use of the Service, if you did not modify your use as advised;
5. During or with respect to preview, pre-release, beta or trial versions of a Service, feature or software (as determined by us) or to purchases made using Microsoft subscription credits;
6. That result from your unauthorized action or lack of action when required, or from your employees, agents, contractors, or vendors, or anyone gaining access to our network by means of your passwords or equipment, or otherwise resulting from your failure to follow appropriate security practices;
7. That result from your failure to adhere to any required configurations, use supported configurations or platforms, follow any policies for acceptable use, or your use of the Service in a manner inconsistent with the features and functionality of the Service (for example, attempts to perform operations that are not supported) or inconsistent with our published guidance;
8. That result from faulty input, instructions, or arguments (for example, requests to access files that do not exist);
9. That result from your attempts to perform operations that exceed prescribed quotas or that resulted from our throttling of suspected abusive behavior or hacked accounts;
10. Due to your use of Service features that are outside of associated Support Windows; or
11. For licenses reserved, but not paid for, at the time of the Incident.
12. Your initiated operations such as restart, stop, start, failover, scale compute, and storage (which by their nature include capacity constraints) that incur downtime are excluded from the uptime calculation unless explicitly defined in the applicable Service Level);
13. Monthly maintenance window that incurs a downtime to patch your server and infrastructure is excluded from the uptime calculation.
14. That result from performance degradation or latency issues without actual service unavailability (except for services which explicitly include a performance-based Service Level).

Microsoft outage communications are intended to help customers take preventive actions for their business-critical applications and are not a confirmation of missed Service Levels or Service Credits eligibility.

Services purchased through Open, Open Value, and Open Value Subscription volume licensing agreements, and Services in an Office 365 Small Business Premium suite purchased in the form of a product key are not eligible for Service Credits based on service fees. For these Services, any Service Credit that you may be eligible for will be credited in the form of service time (i.e., days) as opposed to service fees, and any references to “Applicable Service Fees” is deleted and replaced by “Applicable Period.”

Service Specific Terms

Microsoft Dynamics 365

Dynamics 365 Business Central

Downtime: Any period of time when end users are unable to login to their instance.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Dynamics 365 Commerce

Additional Definitions:
"**Active Tenant**" means a tenant with an active high availability production topology in the Management Portal that (A) has been deployed to a Partner Application Service; and (B) has an active database that users can log into.
"**Partner Application Service**" means a partner application built on top of and combined with the Platform that (A) is used for processing your organization's actual business transactions; and (B) has reserve compute and storage resources equal to or greater than one of the Scale Units your partner selected for the applicable partner application.
"**Maximum Available Minutes**" means the total accumulated minutes during an Applicable Period in which an Active Tenant was deployed in a Partner Application Service using an active high availability production topology.
"**Platform**" means the Service's client forms, SQL server reports, batched operations, and API endpoints, or the Service's retail APIs that are used for commerce or retail purposes only.
"**Scale Unit**" means the increments by which compute and storage resources are added to or removed from a Partner Application Service.
"**Service Infrastructure**" means the authentication, computing, and storage resources that Microsoft provides in connection with the Service.
Downtime: Any period of time when end users are unable to access their Active Tenant, due to a failure in the unexpired Platform or the Service Infrastructure as Microsoft determines from automated health monitoring and system logs. Downtime does not include Scheduled Downtime, the unavailability of Service add-on features, the inability to access the Service due to your modifications of the Service, or periods where the Scale Unit capacity is exceeded.
Uptime Percentage: The Uptime Percentage for a given Active Tenant in an Applicable Period is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Dynamics 365 Contact Center

Downtime: Any period of time when end users are unable to initiate or receive PSTN calls or perform real-time communications (e.g. IVR routing).
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period Downtime is the sum of the length (in minutes) of each incident that occurs during that Applicable Period multiplied by the number of users impacted by that incident. Credit will be paid only against the actual service(s) that are impacted.

This SLA does not include Scheduled Downtime, the unavailability of Service add-on features, the inability to access the Service due to your modifications of the Service, outages caused by any failure of third-party software, equipment, or services that are not controlled by Microsoft, or Microsoft software that is not being run by Microsoft itself as part of the Service.

Service Credit:

Uptime Percentage	Service Credit
< 99.99%	5%
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Dynamics 365 Customer Insights

Downtime: Any period of time when end users are unable to login to their environment. Downtime does not include Scheduled Downtime, the unavailability of Service add-on features, or the inability to access the Service due to your modifications of the Service.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Dynamics 365 Customer Service Enterprise; Dynamics 365 Customer Service Professional; Dynamics 365 Customer Service Insights; Dynamics 365 Field Service; Dynamics 365 Marketing

Downtime: Any period of time when end users are unable to read or write any Service data for which they have appropriate permission but this does not include non-availability of Service add-on features.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Dynamics 365 Guides

Additional Definitions:
Downtime: Any period of time when end user is unable to read or write any Service data for which they have appropriate permission. Any period of time when end users are unable to initiate or participate in calls.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.

* Downtime does not include Scheduled Downtime.

Service Credit:

Uptime Percentage	Service Credit
<99.5%	25%
<99%	50%

Table of Contents / Definitions

Dynamics 365 Human Resources

Additional Definitions:
"**Active Tenant**" means a tenant with an active high availability production topology in the Management Portal that has an active database that users can log into.
Downtime: Any period of time when end users are unable to read or write any Service data for which they have appropriate permission. Downtime does not include Scheduled Downtime.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.

Service Credit:

Uptime Percentage	Service Credit
< 99.5%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Dynamics 365 Intelligent Order Management

Downtime: Any period of time when end users are unable to read or write any Service data for which they have appropriate permission but this does not include any non-availability of Service add-on features. Downtime does not include Scheduled Downtime.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Dynamics 365 Remote Assist

Additional Definitions:
Downtime: Any period of time when end users are unable to conduct instant messaging conversations or initiate or participate in calls.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
**Instant messaging conversations available only in some platforms*

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%

Table of Contents / Definitions

Dynamics 365 Sales Enterprise; Dynamics 365 Sales Professional

Downtime: Any period of time when end users are unable to read or write any Service data for which they have appropriate permission, but this does not include non-availability of Service add-on features.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Dynamics 365 Supply Chain Management; Dynamics 365 Finance; Dynamics 365 Project Operations

Additional Definitions:
"**Active Tenant**" means a tenant with an active high availability production topology in the Management Portal that (A) has been deployed to a Partner Application Service; and (B) has an active database that users can log into.
"**Partner Application Service**" means a partner application built on top of and combined with the Platform that (A) is used for processing your organization's actual business transactions; and (B) has reserve compute and storage resources equal to or greater than one of the Scale Units your partner selected for the applicable partner application.
"**Maximum Available Minutes**" means the total accumulated minutes during an Applicable Period in which an Active Tenant was deployed in a Partner Application Service using an active high availability production topology.
"**Platform**" means the Service's client forms, SQL server reports, batched operations, and API endpoints, or the Service's retail APIs that are used for commerce or retail purposes only.
"**Scale Unit**" means the increments by which compute, and storage resources are added to or removed from a Partner Application Service.
"**Service Infrastructure**" means the authentication, computing, and storage resources that Microsoft provides in connection with the Service.
Downtime: Any period of time when end users are unable to login to their Active Tenant, due to a failure in the unexpired Platform or the Service Infrastructure as Microsoft determines from automated health monitoring and system logs. Downtime does not include Scheduled Downtime, the unavailability of Service add-on features, the inability to access the Service due to your modifications of the Service, or periods where the Scale Unit capacity is exceeded.
Uptime Percentage: The Uptime Percentage for a given Active Tenant in an Applicable Period is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Office 365 Services

Duet Enterprise Online

Downtime: Any period of time when users are unable to read or write any portion of a SharePoint Online site collection for which they have appropriate permissions.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Service Level Exceptions: This SLA does not apply when the inability to read or write any portion of a SharePoint Online site is caused by any failure of third-party software, equipment, or services that are not controlled by Microsoft, or Microsoft software that is not being run by Microsoft itself as part of the Service.

Additional Terms: You will be eligible for a Service Credit for Duet Enterprise Online only when you are eligible for a Service Credit for the SharePoint Online Plan 2 User SLs that you have purchased as a prerequisite for your Duet Enterprise Online User SLs.

Table of Contents / Definitions

Exchange Online

Downtime: Any period of time when users are unable to send or receive email with Outlook Web Access. There is no Scheduled Downtime for this service.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

OWA Availability service level does not apply to the following scenarios:

- Denial of Service attacks (DoS)D
- Microsoft 365 Tenant misconfiguration
- Network issues outside Microsoft 365 Boundary
- Exceeding Microsoft 365 send/receive limits
- Issues due to Extensions (i.e tenant custom policies, apps)
- Third Party induced incidents (i.e, ISP, OnPrem, etc)

Uptime Percentage for Exchange Email Delivery Time
M365 Internal Email Delivery
M365 Internal Email Delivery is defined as the fastest 95% of messages measured in seconds over an Applicable Period within the Microsoft 365 boundary and applies to the following scenarios:

- **Inbound to Microsoft 365 cloud hosted mailbox**: The elapsed time from when a mail enters the Microsoft 365 boundary to when the mail is delivered to the Microsoft 365 cloud hosted mailbox.
- **Intra-tenant Microsoft 365 to Microsoft 365 cloud hosted mailbox (Inter-Tenant Excluded)**: The elapsed time from when a Microsoft 365 cloud hosted mailbox sends mail, and that mail is delivered to another Microsoft 365 cloud hosted mailbox.
- **Microsoft 365 cloud hosted mailbox to an external recipient**: The elapsed time from when a Microsoft 365 cloud hosted mailbox sends mail to an external recipient, where the delay has been identified as a failure within the Microsoft 365 tenant boundary.
- **Microsoft 365 relay to an external recipient**: The elapsed time from when a mail enters the Microsoft 365 boundary from a customer's on-premises environment via an inbound connector and leaves the Microsoft 365 tenant boundary, and the delay has been identified as a failure within the Microsoft 365 boundary.

M365 Internal Email Delivery is measured then sorted by elapsed time. The fastest 95% of measurements are used to create the average delivery time for the Applicable Period.

Customers are eligible for a service credit when the fastest 95% of email delivery for the Applicable Period exceeds the following thresholds:

Average Email Delivery Time	Service Credit
>1 minute	25%
>4 minutes	50%
>10 minutes	100%

Guaranteed Email Delivery

Guaranteed Email Delivery within the Microsoft 365 boundary refers to the successful delivery of email messages. There is no scheduled downtime for this and applies only to failures within the M365 boundary.

Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{Total\,minutes \in a\,month - Downtime}{Total\,minutes \in a\,month} x100$$

where Downtime is the sum of the time (in minutes) multiplied by the fraction of the service that is not available summed over the calendar month.

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

M365 Email Ingress/Egress Availability

M365 Email Ingress/Egress Availability is defined as any period of time M365 is not able to accept or send email messages due to a problem within the M365 boundary. There is no scheduled downtime for this.

Availability is divided into two categories:

1. Permanent rejections due to M365 issues
2. Transient rejections due to M365 issues

Uptime Percentage

1. Permanent rejections due to M365 issues

The Uptime Percentage is calculated using the following formula:

$$\frac{Total\,minutes \in a\,month - Downtime}{Total\,minutes \in a\,month} x100$$

where Downtime is the sum of the time (in minutes) multiplied by the fraction of the service that is not available summed over the calendar month.

2. Transient rejections due to M365 issues

Uptime Percentage

$$\frac{Total\,messages \in a\,month - Impacted\,Messages}{Total\,messages\,for\,the\,month} x100$$

Impacted Messages are defined as messages delayed more than (10) minutes during the issues summed over the calendar month.

Service Credit

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

The Email Delivery Service Level applies only to legitimate business email delivered to or sent by valid Microsoft 365 licensed email accounts. This Email Delivery Service Level does not apply to:

- Customer induced mail storms
- Bulk email (customer mailers, newsletters, etc.)
- Delivery of email to an archive
- Denial of Service attacks (DoS)
- Microsoft 365 Tenant misconfiguration
- Email delays where the failure mode is within the customers on-premises boundary and/or third-party service providers.
- Network latency between Microsoft 365 and end user email clients
- Mail that is throttled by Microsoft 365 to protect the health of the service or because a tenant has exceeded defined send and/or receive messaging limits.

- Messages de-prioritized by Microsoft 365 to protect the overall health of the service (e.g., large messages, messages sent to a large number of recipients or distribution lists that contain a substantial number of recipients).
- Messages delayed due to expensive customers configured rules and policies.

Additional Terms: See Appendix 1 – Service Level Commitment for Virus Detection and Blocking, Spam Effectiveness, or False Positive.

Table of Contents / Definitions

Exchange Online Archiving

Downtime: Any period of time when users are unable to access the email messages stored in their archive. There is no Scheduled Downtime for this service.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Service Level Exceptions: This SLA does not apply to the Enterprise CAL suite purchased through Open Value and Open Value Subscription volume licensing agreements.

Table of Contents / Definitions

Exchange Online Protection

Downtime: Any period of time Microsoft 365 is not able to accept or send email messages due to a problem within the Microsoft 365 boundary. There is no Scheduled Downtime for this service.

Exchange Online Protection Availability is divided into two categories:
1. Permanent rejections due to Microsoft 365 issues
2. Transient rejections due to Microsoft 365 issues

Uptime Percentage
1. Permanent rejections due to Microsoft 365 issues.

The Uptime Percentage is calculated using the following formula:

$$\frac{Total\,minutes \in a\,month - Downtime}{Total\,minutes \in a\,month} x100$$

where Downtime is duration of the issue (in minutes) multiplied by the fraction of the service that is not available summed over the calendar month.
2. Transient rejections due to Microsoft 365 issues

Uptime Percentage

$$\frac{Total\,messages\,for\,the\,month - Impacted\,Messages}{Total\,Messages\,for\,the\,month} x100$$

Impacted Messages are defined as messages delayed more than (10) minutes during the issue summed over the calendar month.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Exchange Online Protection applies only to legitimate business email delivered to or sent by valid Microsoft 365 licensed email accounts. This does not apply to:
- Customer induced mail storms
- Bulk email (customer mailers, newsletters, etc.)
- Delivery of email to an archive
- Denial of Service attacks (DoS)

- Microsoft 365 Tenant misconfiguration
- Email delays where the failure mode is within the customers on-premises boundary and/or third-party service providers.
- Network latency between Microsoft 365 and end user email clients
- Mail that is throttled by Microsoft 365 to protect the health of the service or because a tenant has exceeded defined send and/or receive messaging limits.
- Messages de-prioritized by Microsoft 365 to protect the overall health of the service (For example: large messages, messages sent to a large number of recipients or distribution lists that contains substantial number of recipients)
- Messages delayed due to expensive customer configured rules and policies

Additional Terms: See (i) Appendix 1 – Service Level Commitment for Virus Detection and Blocking, Spam Effectiveness, or False Positive.

Table of Contents / Definitions

Microsoft MyAnalytics

Downtime: Any period of time when users are unable to access the MyAnalytics dashboard.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Microsoft Stream (Classic)

Downtime: Any period of time when users are unable to upload, playback, delete video or edit video metadata when they have appropriate permissions and content is valid excluding unsupported scenarios.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Level Commitment:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

[1]Unsupported Scenarios could include playback on unsupported devices / OS, client-side network issues, and user errors.

Table of Contents / Definitions

Microsoft Teams

Downtime: Any period of time when end users are unable to conduct instant messaging conversations or initiate online meetings.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Microsoft 365 Apps for business

Downtime: Any period of time when Office applications are put into reduced functionality mode due to an issue with Office 365 activation.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Microsoft 365 Apps for enterprise

Downtime: Any period of time when Office applications are put into reduced functionality mode due to an issue with Office 365 activation.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Office 365 Advanced Compliance

Downtime: Any period of time when Customer Lockbox component of Office 365 Advanced Compliance is put into reduced functionality mode due to an issue with Office 365.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Office Online

Downtime: Any period of time when users are unable to use the Web Applications to view and edit any Office document stored on a SharePoint Online site for which they have appropriate permissions.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Office 365 Video

Downtime: Any period of time when users are unable to upload, view or edit videos in the video portal when they have appropriate permissions and valid content.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Level Commitment:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

OneDrive for work or school

Downtime: Any period of time when users are unable to view or edit files stored on their OneDrive (work/school) storage.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Project

Downtime: Any period of time when users are unable to read or write any portion of a SharePoint Online site collection with Project Web App for which they have appropriate permissions.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

SharePoint Online

Downtime: Any period of time when users are unable to read or write any portion of a SharePoint Online site collection for which they have appropriate permissions.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Microsoft Teams – Calling Plans, Teams Phone and Audio Conferencing

Downtime: Any period of time when end users are unable to initiate a PSTN call or unable to dial into conference audio via the PSTN, or process calls with Call Queues or Auto Attendant.
Uptime Percentage: The Uptime Percentage is calculated using the following formula for each of the services:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period Downtime is the sum of the length (in minutes) of each incident that occurs during that Applicable Period multiplied by the number of users impacted by that incident. Credit will be paid only against the actual service(s) that are impacted.

This SLA does not apply to outages caused by any failure of third-party software, equipment, or services that are not controlled by Microsoft, or Microsoft software that is not being run by Microsoft itself as part of the Service.
Service Credit:

Uptime Percentage	Service Credit
< 99.999%	5%
< 99.99%	10%
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Viva Engage

Downtime: Any period of time greater than ten minutes when more than five percent of end users are unable to post or read messages on any portion of the Viva Engage network for which they have appropriate permissions.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Microsoft Teams – Voice Quality

This SLA applies to any eligible call placed by any voice service user within the subscription (enabled for making any type of call VOIP or PSTN).
Additional Definitions:
"**Eligible Call**" is a Microsoft Teams placed call (within a subscription) that meets both conditions below:

- The call was placed from a Microsoft Teams Certified IP Desk phones on wired Ethernet.
- Packet Loss, Jitter and Latency issues on the call were due to networks managed by Microsoft.

"**Total Calls**" is the total number of Eligible Calls.
"**Poor Quality Calls**" is the total number of Eligible Calls that are classified as poor based on numerous factors that could impact call quality in the networks managed by Microsoft. While the current Poor Call classifier is built primarily on network parameters like RTT (Roundtrip Time), Packet Loss Rate, Jitter and Packet Loss-Delay Concealment Factors, it is dynamic and continually updated based on new learnings from analysis using millions of Skype, Skype for Business, and Microsoft Teams calls and evolution of Devices, Algorithms and end user ratings.
Good Call Rate: The Good Call Rate is calculated using the following formula:

$$\frac{Total\,Calls - Poor\,Quality\,Calls}{Total\,Calls} x100$$

Service Credit:

Good Call Rate	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Workplace Analytics

Downtime: Any period of time when users are unable to access the Workplace Analytics website.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Microsoft Azure Services and Plans

Microsoft Entra ID

Microsoft Entra ID Basic and Microsoft Entra ID Premium
Additional Definitions:
Downtime: Any period of time when users are unable to log in to the Microsoft Entra ID service, or Microsoft Entra ID fails to successfully emit the authentication and authorization tokens required for users to log into applications connected to the service.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.99%	10%
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Azure Active Directory B2C

Additional Definitions:
"**Deployment Minutes**" is the total number of minutes for which an Azure AD B2C directory has been deployed during an Applicable Period.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Azure AD B2C directories in a given Microsoft Azure subscription during an Applicable Period.
Downtime: is the total accumulated minutes across all Azure AD B2C directories deployed by Customer in a given Microsoft Azure subscription during which the Azure AD B2C service is unavailable. A minute is considered unavailable if all attempts to process user sign-up and sign-in fail to return tokens or valid Error Codes, or do not return responses within two minutes.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{Maximum\,Available\,Minutes - Downtime}{Maximum\,Available\,Minutes} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.99%	10%
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Microsoft Entra Domain Services

Additional Definitions:
"**Managed Domain**" refers to an Active Directory domain that is provisioned and managed by Microsoft Entra Domain Services.
"**Maximum Available Minutes**" is the total number of minutes that a given Managed Domain has been deployed by Customer in Microsoft Azure during an Applicable Period in a given Microsoft Azure subscription.
"**Downtime**" is the total accumulated minutes during an Applicable Period for a given Microsoft Azure subscription during which a given Managed Domain is unavailable. A minute is considered unavailable if all requests for domain authentication of user accounts belonging to the Managed Domain, LDAP bind to the root DSE, or DNS lookup of records, made from within the virtual network where the Managed Domain is enabled, either return an Error Code or fail to return a Success Code within 30 seconds.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Levels and Service Credits are applicable to Customer's use of Microsoft Entra Domain Services:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Analysis Services

Additional Definitions:
"**Server**" means any Azure Analysis Services server.
"**Maximum Available Minutes**" is the total number of minutes that a given Server has been deployed in Microsoft Azure during an Applicable Period in a given Microsoft Azure subscription.
"**Client Operations**" is the set of all documented operations supported by Azure Analysis Services.
Downtime: is the total accumulated minutes during an Applicable Period for a given Microsoft Azure subscription during which a given Server is unavailable. A minute is considered unavailable for a given Server if more than 1% of all Client Operations completed during the minute return an Error Code.
Uptime Percentage: The Uptime Percentage for a given Server is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure API for FHIR

"**Total Transaction Attempts**" is the total number of authenticated Azure API for FHIR requests made by Customer during an Applicable Period for a given Microsoft Azure subscription. Total Transaction Attempts do not include API requests that return an Error Code that are continuously repeated within a five-minute window after the first Error Code is received.
"**Failed Transactions**" is the set of all requests within Total Transaction Attempts that result in an Error Code or otherwise do not return a Success Code within 60 seconds after receipt by the Azure API for FHIR Service.
Uptime Calculation
"Uptime Percentage" for Azure API for FHIR is calculated as Total Transaction Attempts less Failed Transactions divided by Total Transaction Attempts multiplied by 100. Uptime Percentage is represented by the following formula:

$$\frac{Total\ Transaction\ Attempts - Failed\ Transactions}{Total\ Transaction\ Attempts} x100$$

The following Service Levels and Service Credits are applicable to Azure API for FHIR:
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

API Center Services

Additional Definitions:
Deployment Minutes: is the total number of minutes for which an API Center instance has been purchased during an Applicable Period.
Assets: means any API Center asset registration in an API Center instance (such as Apis, Api Definitions, Api Versions, Api Deployments, Environments, Metadata).
Maximum Available Minutes: is the sum of all Deployment Minutes for an API Center instance associated with a given Microsoft Azure subscription during an Applicable Period.
Downtime: is the total accumulated minutes, during which a given API Center data plane is unavailable. A minute is considered unavailable for the given API Center instance if all attempts by users to execute API calls against API Center data plane result in an Error Code or do not return a response within five minutes.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{Maximum\,Available Minutes - Downtime}{Maximum\,Available Minutes} x100$$

Service Credit for Free Plan scaled within a single region: No SLA
Service Credit for Standard Plan scaled within a single region:

Uptime Percentage	Service Credit
< 99.95%	10%
< 99%	25%

Table of Contents / Definitions

API Management Services

Additional Definitions:
"**Deployment Minutes**" is the total number of minutes that a given API Management instance has been deployed in Microsoft Azure during an Applicable Period.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all API Management instances deployed by you in a given Microsoft Azure subscription during an Applicable Period.
"**Proxy**" is the component of the API Management Service responsible for receiving API requests and forwarding them to the configured dependent API.
Downtime: The total accumulated Deployment Minutes, across all API Management instances deployed by you in a given Microsoft Azure subscription, during which the API Management Service is unavailable. A minute is considered unavailable for a given API Management instance if all continuous attempts to perform operations through the Proxy throughout the minute result in either an Error Code or do not return a Success Code within five minutes.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit for Consumption Tier, Basic Tier, Standard Tier and Premium Tier deployments scaled within a single region:

Uptime Percentage	Service Credit
< 99.95%	10%
< 99%	25%

Service Credit for Premium Tier deployments scaled across two or more regions:

Uptime Percentage	Service Credit
< 99.99%	10%
< 99%	25%

Table of Contents / Definitions

App Center

Additional Definitions:
"**Build Service**" is a feature that allows customers to build their mobile applications in Visual Studio App Center.
"**Test Service**" is a feature that allows customers to upload and run tests for their mobile applications on physical devices running in Visual Studio App Center.
"**Push Notification Service**" is a feature that enables customers to push messages to specific devices configured to receive such messages using Visual Studio App Center.

Uptime Calculation and Service Levels for Visual Studio App Center Build Service
"**Maximum Available Minutes**" is the total number of minutes for which Build Service has been deployed by Customer for a given Microsoft Azure subscription during an Applicable Period.
"**Downtime**" is the total number of minutes within Maximum Available Minutes during which the Build Service is unavailable. A minute is considered unavailable if all continuous HTTP requests to the Build Service to perform operations initiated by Customer throughout the minute either result in an Error Code or do not return a response within one minute.
Uptime Percentage: The Uptime Percentage for the Visual Studio App Center Build Service is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes multiplied by 100. Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Visual Studio App Center Build Service.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Uptime Calculation and Service Levels for Visual Studio App Center Test Service
"**Maximum Available Minutes**" is the total number of minutes for which Test Service has been deployed by Customer for a given Microsoft Azure subscription during an Applicable Period.
Downtime: The total number of minutes within Maximum Available Minutes during which the Test Service is unavailable. A minute is considered unavailable if all continuous HTTP requests to the Test Service to perform operations initiated by Customer throughout the minute either result in an Error Code or do not return a response within one minute.
Uptime Percentage: The Uptime Percentage for the Visual Studio App Center Test Service is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes multiplied by 100. Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Visual Studio App Center Test Service.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Uptime Calculation and Service Levels for Visual Studio App Center Push Notification Service
"**Maximum Available Minutes**" is the total number of minutes for which Push Notification Service has been deployed by Customer for a given Microsoft Azure subscription during an Applicable Period.
Downtime: The total number of minutes within Maximum Available Minutes during which Push Notification Service is unavailable. A minute is considered unavailable if all continuous HTTP requests to Push Notification Service to perform operations initiated by Customer throughout the minute either result in an Error Code or do not return a response within one minute.
Uptime Percentage: The Uptime Percentage for the Visual Studio App Center Push Notification Service is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes multiplied by 100. Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Visual Studio App Center Push Notification Service.

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

App Configuration

Additional Definitions

"**Configuration Store**" refers to a single deployment of Azure App Configuration created by Customer, such that it is enumerated in the App Configuration tab in the Management Portal.

Uptime Calculation and Service Levels for Azure App Configuration

"**Deployment Minutes**" is the total number of minutes that a given Configuration Store has been deployed in Microsoft Azure during an Applicable Period.

"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Configuration Stores deployed by Customer in a given Microsoft Azure subscription during an Applicable Period.

"**Downtime**" is the total accumulated minutes within Maximum Available Minutes, during which the Configuration Store is unavailable. A minute is considered unavailable for a given Configuration Store when there is no connectivity throughout the minute between the Configuration Store and Microsoft's Internet gateway.

"**Uptime Percentage**" for Azure App Configuration is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for a given Microsoft Azure subscription. Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of App Configuration.

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

App Service

Additional Definitions:

"**Availability Zone**" is a fault-isolated area within an Azure region, providing redundant power, cooling, and networking.

"**Deployment Minutes**" is the total number of minutes that a given App has been set to running in Microsoft Azure during an Applicable Period. Deployment Minutes is measured from when the App was created or the Customer initiated an action that would result in running the App to the time the Customer initiated an action that would result in stopping or deleting the App.

"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Apps deployed by Customer in a given Microsoft Azure subscription during an Applicable Period

"**App**" is a Web App, Mobile App, API App, Logic App, deployed by Customer within the App Service. The SLA is supported when running on a single instance and on multiple instances.

Uptime Calculation and Service Levels for App Service Apps in Availability Zones

"**Maximum Available Minutes**" is the total accumulated Deployment Minutes across all Availability Zone enabled Apps deployed by Customer in a given Microsoft Azure subscription during an Applicable Period.

"Downtime" is the total accumulated Deployment Minutes, across all Availability Zone enabled Apps deployed by Customer in a given Microsoft Azure subscription, during which the App is unavailable. A minute is considered unavailable for a given Availability Zone enabled App when there is no connectivity between the App and Microsoft's Internet gateway.

"**Uptime Percentage**" is calculated using the following formula:

$$\frac{\text{(Maximum Available Minutes-Downtime)}}{\text{Maximum Available Minutes}} x100$$

Service Credit:
The following Service Levels and Service Credits are applicable to Customer's use of Apps deployed across two or more Availability Zones in the same region:

Uptime Percentage	Service Credit
< 99.99%	10%
< 99%	25%
< 95%	100%

Uptime Calculation and Service Levels for App Service Apps that don't use Availability Zones
"**Maximum Available Minutes**" is the total accumulated Deployment Minutes across all Apps deployed by Customer in a given Microsoft Azure subscription an Applicable Period.
"**Downtime**" is the total accumulated Deployment Minutes, across all Apps deployed by Customer in a given Microsoft Azure subscription, during which the App is unavailable. A minute is considered unavailable for a given App when there is no connectivity between the App and Microsoft's Internet gateway.
"**Uptime Percentage**" is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit:
The following Service Levels and Service Credits are applicable to Customer's use of Apps that don't use Availability Zones:

Uptime Percentage	Service Credit
< 99.95%	10%
< 99%	25%
< 95%	100%

Additional Terms: Service Credits are applicable only to fees attributable to your use of Web Apps, Mobile Apps, API Apps or Logic Apps and not to fees attributable to other types of apps available through the App Service, which are not covered by this SLA.

As of September 1, 2024, App Service Environment version 1 and version 2 are retired Service features, and their Support Windows are expired. App Service no longer provides any Service Level guarantees, and by extension Service Credits, for performance or availability issues related to workloads running on App Service Environment version 1 and version 2.

Table of Contents / Definitions

Application Gateway

Additional Definitions:
"**Application Gateway Cloud Service**" refers to a collection of two or more medium or larger Application Gateway instances or deployments capable of supporting auto scale or zone redundancy, configured to perform HTTP load balancing services.
"**Maximum Available Minutes**" is the total accumulated minutes during an Applicable Period during which an Application Gateway Cloud Service has been deployed in a Microsoft Azure subscription.
Downtime: is the total accumulated Maximum Available Minutes during an Applicable Period for a given Application Gateway Cloud Service during which the Application Gateway Cloud Service is unavailable. A given minute is considered unavailable if all attempts to connect to the Application Gateway Cloud Service throughout the minute are unsuccessful.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.95%	10%
< 99%	25%

Table of Contents / Definitions

Application Gateway for Containers

Additional Definitions:
"**Application Gateway for Containers**" refers to both control plane and data plane concepts to perform HTTP load balancing services.
"**ALB Controller**" refers to the component installed into a customer's Kubernetes cluster that is responsible for translating and transmitting user defined configuration in the Kubernetes cluster to Application Gateway for Containers.
"**Maximum Available Minutes**" is the total accumulated minutes during a billing month during which an Application Gateway for Containers Service has been deployed in an Azure subscription.
"**Control Plane Downtime**" is the total accumulated minutes during a billing month for a given Application Gateway for Containers resource during which changes to the Application Gateway for Containers control plane are unavailable. A given minute is considered unavailable if all connections initiated by Application Gateway for Container's ALB Controller throughout the minute are unsuccessful.
"**Data Plane Downtime**" is the total accumulated minutes during a billing month for a given Application Gateway for Containers deployment during which the Application Gateway for Containers data plane is unavailable. A given minute is considered unavailable if all attempts to connect to a frontend of Application Gateway for Containers throughout the minute are unsuccessful.
Monthly Uptime Percentage: The Monthly Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Control Plane Downtime-Data Plane Downtime}}{\text{Maximum Available Minutes}}$$

Service Credit:

Uptime Percentage	Service Credit
< 99.95%	10%
< 99%	25%

Table of Contents / Definitions

Application Insights

Additional Definitions:
"**Application Insights Resource**" is the container in Application Insights that collects, processes and stores the data for a single instrumentation key.
"**Maximum Available Minutes**" is the total number of minutes that a given Application Insights Resource has been deployed by Customer within a Microsoft Azure subscription during an Applicable Period.
"**Downtime**" is the total number of minutes within Maximum Available Minutes that data within an Application Insights Resource are unavailable. A minute is considered unavailable for a given Application Insights Resource during which no HTTP operations resulted in a Success Code.
Query Availability Percentage: for a given Application Insights Resource in an Applicable Period is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes multiplied by 100.
The Query Availability Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Application Insights Service – Query Availability SLA:

Query Availability Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure Arc

Additional Definitions

"**Maximum Available Minutes**" is the total accumulated minutes in an Applicable Period during which at least one Kubernetes configuration Azure resource has been deployed on an Azure Arc enabled Kubernetes resource in a Microsoft Azure subscription.
"**Downtime**" is the total accumulated Maximum Available Minutes in an Applicable Period during which at least one Kubernetes configuration Azure resource has been deployed on an Azure Arc enabled Kubernetes resource, but the REST API operations for the Kubernetes configuration Azure resource are unavailable.
"**Uptime Percentage**" The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of Kubernetes configuration Azure resource on top of Azure Arc enabled Kubernetes:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Automation

Automation Service – Desired State Configuration (DSC)
Additional Definitions:
"**Deployment Minutes**" is the total number of minutes that a given Automation account has been deployed in Microsoft Azure during an Applicable Period.
"**DSC Agent Service**" is the component of the Automation Service responsible for receiving and responding to pull, registration, and reporting requests from DSC nodes.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Automation accounts deployed in a given Microsoft Azure subscription during an Applicable Period
Downtime: The total accumulated Deployment Minutes, across all Automation accounts deployed in a given Microsoft Azure subscription, during which the DSC Agent Service is unavailable. A minute is considered unavailable for a given Automation account if all continuous pull, registration, and reporting requests from DSC nodes associated with the Automation account to the DSC Agent Service throughout the minute either result in an Error Code or do not return a Success Code within five minutes.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Additional Terms: Service Credits are applicable only to fees attributable to your use of DSC functionality within the Automation Service.
Automation Service – Process Automation
Additional Definitions:
"**Delayed Jobs**" is the total number of Jobs, for a given Microsoft Azure subscription, that fail to start within thirty (30) minutes of their Planned Start Times.
"**Job**" means the execution of a Runbook.
"**Planned Start Time**" is a time at which a Job is scheduled to begin executing.
"**Runbook**" means a set of actions specified by you to execute within Microsoft Azure.
"**Total Jobs**" is the total number of Jobs scheduled for execution during a given Applicable Period, for a given Microsoft Azure subscription.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{Total\ Jobs - Delayed\ Jobs}{Total\ Jobs} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Additional Terms: Service Credits are applicable only to fees attributable to your use of Process Automation functionality within the Automation Service.

Table of Contents / Definitions

Azure Backup

Additional Definitions:
"**Backup**" or "**Back Up**" is the process of copying computer data from a registered server to a Backup Vault.
"**Backup Agent**" refers to the software installed on a registered server that enables the registered server to Back Up or Restore one or more Protected Items.
"**Backup Vault**" refers to a container in which you may register one or more Protected Items for Backup.
"**Failure**" means that either the Backup Agent or the Service fails to fully complete a properly configured Backup or Recovery operation due to unavailability of the Backup Service.
"**Protected Item**" refers to a collection of data, such as a volume, database, or virtual machine that has been scheduled for Backup to the Backup Service such that it is enumerated as a Protected Item in the Protected Items tab in the Recovery Services section of the Management Portal.
"**Recovery**" or "**Restore**" is the process of restoring computer data from a Backup Vault to a registered server.
Uptime Calculation and Service Levels for Backup Service
Additional Definitions:
"**Deployment Minutes**" is the total number of minutes during which a Protected Item has been scheduled for Backup to a Backup Vault.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Protected Items for a given Microsoft Azure subscription during an Applicable Period.
Downtime: The total accumulated Deployment Minutes across all Protected Items scheduled for Backup by you in a given Microsoft Azure subscription during which the Backup Service is unavailable for the Protected Item. The Backup Service is considered unavailable for a given Protected Item from the first Failure to Back Up or Restore the Protected Item until the initiation of a successful Backup or Recovery of a Protected Item, provided that retries are continually attempted no less frequently than once every thirty minutes.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure Bastion

Additional Definitions
Uptime Calculation
"**Maximum Available Minutes**" is the total accumulated minutes during an Applicable Period during which a given Azure Bastion has been deployed in a Microsoft Azure subscription.
"**Downtime**" is the total accumulated Maximum Available Minutes during which an Azure Bastion is unavailable. A minute is considered unavailable if all attempts to connect to the Azure Bastion within the minute are unsuccessful.
"**Uptime Percentage**" for a given Azure Bastion is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for a given Microsoft Azure subscription. Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of each Azure Bastion:

Uptime Percentage	Service Credit
< 99.95%	10%
< 99%	25%

Table of Contents / Definitions

Batch

Additional Definitions:
"**Average Error Rate**" for an Applicable Period is the sum of Error Rates for each hour in the billing Applicable Period divided by the total number of hours in the Applicable Period.
"**Error Rate**" is the total number of Failed Requests divided by Total Requests during a given one-hour interval. If the Total Requests in a given one-hour interval is zero, the Error Rate for that interval is 0%.
"**Excluded Requests**" are requests that result in an HTTP 4xx status code, other than an HTTP 408 status code.
"**Failed Requests**" is the set of all requests within Total Requests that either return an Error Code or an HTTP 408 status code or fail to return a Success Code within 5 seconds.
"**Total Requests**" is the total number of authenticated REST API requests, other than Excluded Requests, to perform operations against Batch accounts attempted within a one-hour interval within a given Azure subscription during an Applicable Period.
Uptime Percentage: for the Batch Service is calculated by subtracting from 100% the Average Error Rate for a given Microsoft Azure subscription in an Applicable Period. The "Average Error Rate" for an Applicable Period is the sum of Error Rates for each hour in the Applicable Period divided by the total number of hours in the Applicable Period. Uptime Percentage is represented by the following formula:

$$\text{Uptime \%} = 100\% - \text{Average Error Rate}$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

BizTalk Services

Additional Definitions:
"**BizTalk Service Environment**" refers to a deployment of the BizTalk Services created by you, as represented in the Management Portal, to which you may send runtime message requests.
"**Deployment Minutes**" is the total number of minutes that a given BizTalk Service Environment has been deployed in Microsoft Azure during an Applicable Period.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all BizTalk Service Environments deployed by you in a given Microsoft Azure subscription during an Applicable Period.
"**Monitoring Storage Account**" refers to the Azure Storage account used by the BizTalk Services to store monitoring information related to the execution of the BizTalk Services.
Downtime: The total accumulated Deployment Minutes, across all BizTalk Service Environments deployed by you in a given Microsoft Azure subscription, during which the BizTalk Service Environment is unavailable. A minute is considered unavailable for a given BizTalk Service Environment when there is no connectivity between your BizTalk Service Environment and Microsoft's Internet gateway.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Service Level Exceptions: The Service Levels and Service Credits are applicable to your use of the Basic, Standard, and Premium tiers of the BizTalk Services. The Developer tier of the Microsoft Azure BizTalk Services is not covered by this SLA.

Additional Terms: When submitting a claim, you must ensure that complete monitoring data is maintained within the Monitoring Storage Account and is made available to Microsoft.

Table of Contents / Definitions

Azure Bot Service

Additional Definitions:
"**Azure Bot Service Premium Channel**" is a Bot Framework channel in the premium category.
"**Bot**" is the developer's Internet facing conversational application which is registered with and is configured to send and receive messages from the Azure Bot Service.
"**Bot Framework**" is a platform for building, connecting, testing, and deploying powerful and intelligent bots.
"**Client**" is the end user facing portion of a Bot.
"**Premium Channels API Endpoint**" is a Bot Framework REST API endpoint for Azure Bot Service Premium Channels

Uptime Calculation and Service Levels for Azure Bot Services Premium Channels:
"**Total API Requests**" is the total number of requests made by the Bot or the Client to the Premium Channel's API Endpoint in a Microsoft Azure subscription during an Applicable Period.
"**Failed API Requests**" are the total number of requests within Total API Requests that return an Error Code or do not respond within 2 minutes.
"**Uptime Percentage**" is calculated as Total API Requests less Failed API Requests divided by Total API Requests multiplied by 100.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Total API Requests-Failed API Requests}}{\text{Total API Requests}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Azure Bot Service Premium Channels.
Service Levels and Service Credits:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure Cache for Redis

Additional Definitions:
"**Cache**" refers to a deployment of the Cache Service created by Customer, such that its Cache Endpoints are enumerated in the Cache tab in the Management Portal.
"**Cache Endpoints**" refers to endpoints through which a Cache may be accessed.
"**Availability Zone**" is a fault-isolated area within an Azure region, providing redundant power, cooling, and networking.

Uptime Calculation and Service Levels for Cache Service
"**Deployment Minutes**" is the total number of minutes that a given Cache has been deployed in Microsoft Azure during an Applicable Period.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Caches deployed by Customer in a given Microsoft Azure subscription during an Applicable Period.
Downtime: The total accumulated Deployment Minutes, across all Caches deployed by Customer in a given Microsoft Azure subscription, during which the Cache is unavailable. A minute is considered unavailable for a given Cache when there is no connectivity throughout the minute between one or more Cache Endpoints associated with the Cache and Microsoft's Internet gateway.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Levels and Service Credits applicable to Customer's use of the Cache Service vary based on the Cache Service's deployment conditions and tier. Unless otherwise provided for above, the Service Levels and Service Credits are applicable to Customer's use of the Cache Service, which includes the Azure Managed Cache Service or the

Standard, Premium, Enterprise, and Enterprise Flash tiers of the Azure Cache for Redis Service. The Basic tier of the Azure Cache for Redis Cache is not covered by this SLA.

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

For any Enterprise or Enterprise Flash tier Cache deployed to three or more Availability Zones in the same Azure region, the following Service Levels and Service Credits are applicable to Customer's use of the Cache Service:

Uptime Percentage	Service Credit
< 99.99%	10%
< 99%	25%

For any Enterprise and Enterprise Flash tier Cache deployed (1) to at least three Azure regions and three or more Availability Zones in each of those regions and (2) with active geo-replication enabled for all Cache instances when the active geo-replication feature is enabled and generally available (i.e., not in preview), the following Service Levels and Service Credits are applicable to Customer's use of the Cache Service:

Uptime Percentage	Service Credit
< 99.999%	10%
< 99%	25%

Table of Contents / Definitions

Azure Managed Redis

Additional Definitions:

“Azure Managed Redis instance” is a deployment of the Azure Managed Redis Service, such that its read/write access endpoints are enumerated in the Management Portal.

“Azure Managed Redis endpoints” refers to endpoints through which an Azure Managed Redis may be accessed.

"High Availability" in the context of Azure Managed Redis means the Azure Managed Redis instance was deployed in a high availability configuration with primary and replica shards across two or more nodes.

“Active Geo-Replication” refers to enabling a geo-replication group with multiple geographically distributed Azure Managed Redis instances to synchronize data in an active-active configuration.

"Availability Zone" is a fault-isolated area within an Azure region, providing redundant power, cooling, and networking.

"Zone Redundancy" refers to an Azure Managed Redis instance deployed using high availability where primary and replica shards are distributed across two or more availability zones.

Uptime Calculation and Service Levels for Azure Managed Redis Service Without Active Geo-Replication

“Maximum Available Minutes” is the total number of minutes for a given Azure Managed Redis instance deployed by Customer in a Microsoft Azure subscription during an Applicable Period.

“Downtime” is the total number of minutes within Maximum Available Minutes during which the Azure Managed Redis instance is unavailable. A minute is considered unavailable for a given Azure Managed Redis instance when there is no connectivity throughout the minute between all Azure Managed Redis Endpoints associated with the Azure Managed Redis instance and Microsoft’s Internet gateway, and Customer has not initiated any product features or management actions that are documented to produce temporary unavailability.

Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and service Credits are applicable to Customer’s use of Memory Optimized, Balanced, Compute Optimized, Flash Optimized tiers of Azure Managed Redis configured in High Availability mode without using Zone Redundancy:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

The following Service Levels and service Credits are applicable to Customer's use of Memory Optimized, Balanced, Compute Optimized, Flash Optimized tiers of Azure Managed Redis configured in High Availability mode with using Zone Redundancy:

Uptime Percentage	Service Credit
< 99.99%	10%
< 99%	25%

Uptime Calculation and Service Levels for Azure Managed Redis Service with Active Geo-Replication
"Maximum Available Minutes" is the total number of minutes for a given Azure Managed Redis instance deployed by Customer in a Microsoft Azure subscription during an Applicable Period.
"Geo-Replication Group" is a logical grouping configuration of Azure Managed Redis instances which synchronize data with each other.
"Downtime" is the total number of minutes within Maximum Available Minutes during which none of the Azure Managed Redis endpoints associated with the geo-replication group are available. A minute is considered unavailable for a given geo-replication group when there is no connectivity throughout the minute between any of the Azure Managed Redis Endpoints for Azure Managed Redis instances associated with the geo-replication group and Microsoft's Internet gateway, and Customer has not initiated any product features or management actions that are documented to produce temporary unavailability.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and service Credits are applicable to Customer's use of Memory Optimized, Balanced, Compute Optimized tiers of Azure Managed Redis that support Active Geo-Replication and are configured so that at least three Azure Managed Redis instances are deployed in at least three regions and are connected using Active Geo-Replication, and each deployment is also using High Availability mode and Zone Redundancy:

Uptime Percentage	Service Credit
< 99.999%	10%
< 99%	25%

Table of Contents / Definitions

Azure Chaos Studio

Additional Definitions:
"Action Minutes" is the total number of minutes that a given Azure Chaos Studio experiment applies an action against a target resource in Microsoft Azure during an Applicable Period. Action Minutes is measure from when an experiment initiates an action to the time when the action has run for the preconfigured duration or is terminated. An experiment can consist of one or more actions running sequentially or concurrently.
"Total Action Minutes" is the sum of all Action Minutes in a given Microsoft Azure subscription during an Applicable Period.
"Downtime" is evaluated in 1-minute intervals for each action that is active, and is the sum of all Action Minutes in a given Microsoft Azure subscription during an Applicable Period, during which Chaos Studio is unavailable. A minute for a given experiment is considered unavailable if within that 1-minute interval at least one experiment stop request for that experiment processed by Chaos Studio returns a 500 error. If no experiment stop request is processed by Chaos Studio within a given 1-minute interval, the Downtime for the interval is assumed to be 0 minute.
"Uptime Percentage": The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Total Action Minutes-Downtime}}{\text{Total Action Minutes}} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure Cloud HSM

Uptime Calculation and Service Levels for Cloud HSM
"**Deployment Minutes**" is the total number of minutes that a given Cloud HSM has been deployed in Microsoft Azure during an Applicable Period.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Cloud HSMs deployed by Customer in a given Microsoft Azure subscription during an Applicable Period.
"**Excluded Transactions**" are transactions for creating, updating, or deleting Cloud HSMs, keys, role assignments, role definitions, and downloading/uploading security domain.
"**Downtime**" is the total accumulated Deployment Minutes, across all Cloud HSMs deployed by Customer in a given Microsoft Azure subscription, during which the Cloud HSM is unavailable. A minute is considered unavailable for a given Cloud HSM if all continuous attempts to perform transactions, other than Excluded Transactions, on the CloudHSM throughout the minute either return an Error Code or do not result in a Success Code within 5 seconds from Microsoft's receipt of the request.
"**Uptime Percentage**" for the Cloud HSM Service is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for a given Microsoft Azure subscription.
Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Cloud HSM Service:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Cloud Services

Additional Definitions:
"**Cloud Services**" refers to a set of compute resources utilized for Web and Worker Roles.
"**Role Instance Connectivity**" is bi-directional network traffic between the role instance and other IP addresses using TCP or UDP network protocols in which the role instance is configured for allowed traffic. The IP addresses can be IP addresses in the same Cloud Service as the virtual machine, IP addresses within the same virtual network as the virtual machine or public, routable IP addresses.
"**Tenant**" represents one or more roles each consisting of one or more role instances that are deployed in a single package.
"**Update Domain**" refers to a set of Microsoft Azure instances to which platform updates are concurrently applied.
"**Web Role**" is a Cloud Services component run in the Azure execution environment that is customized for web application programming as supported by IIS and ASP.NET.
"**Worker Role**" is a Cloud Services component run in the Azure execution environment that is useful for generalized development and may perform background processing for a Web Role.
Uptime Calculation and Service Levels for Cloud Services
"**Maximum Available Minutes**" is the total accumulated minutes during an Applicable Period for all Internet facing roles that have two or more instances deployed in different Update Domains. Maximum Available Minutes is measured from when the Tenant has been deployed and its associated roles have been started resultant from action initiated by Customer to the time Customer has initiated an action that would result in stopping or deleting the Tenant.
Downtime: The total accumulated minutes that are part of Maximum Available Minutes that have no Role Instance Connectivity.
Uptime Percentage: Uptime Percentage is represented by the following formula:

$$Monthly\,Uptime\% = \frac{\text{(Maximum Available Minutes-Downtime)}}{\text{Maximum Available Minutes}} X100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.95%	10%
< 99%	25%

Table of Contents / Definitions

Azure AI Search

Additional Definitions:
"**Average Error Rate**" for an Applicable Period is the sum of Error Rates for each hour in the Applicable Period divided by the total number of hours in the Applicable Period.
"**Error Rate**" is the total number of Failed Requests divided by Total Requests, for a Search Service, during a given one-hour interval. If the Total Requests in a one-hour interval is zero, the Error Rate for that interval is 0%.
"**Excluded Requests**" are all requests that are throttled due to exhaustion of resources allocated for a Search Service Instance, as indicated by an HTTP 503 status code and a response header indicating the request was throttled.
"**Failed Requests**" is the set of all requests within Total Requests that fail to return either a Success Code or HTTP 4xx response.
"**Replica**" is a copy of a search index within a Search Service Instance.
"**Search Service Instance**" is an Azure Search service instance containing one or more search indexes.
"**Total Requests**" is the set of (i) all requests to update a Search Service Instance having three or more Replicas, plus (ii) all requests to query a Search Service Instance having two or more Replicas, other than Excluded Requests, within a one-hour interval during an Applicable Period.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$100\% - Average\,Error\,Rate$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure AI Services

Additional Definitions:
"**Total Transaction Attempts**" is the total number of authenticated API requests by Customer during an Applicable Period for a given Azure AI Service API. Total Transaction Attempts do not include API requests that return an Error Code that are continuously repeated within a five-minute window after the first Error Code is received.
"**Failed Transactions**" is the set of all requests to the Azure AI Service API within Total Transaction Attempts that return an Error Code. Failed Transaction Attempts do not include API requests that return an Error Code that are continuously repeated within a five-minute window after the first Error Code is received.
"**Uptime Percentage**" for each API Service is calculated as Total Transaction Attempts less Failed Transactions divided by Total Transaction Attempts in an Applicable Period for a given API subscription.
Uptime Percentage is represented by the following formula:

$$Monthly\,Uptime\% = \frac{(Total\,Transaction\,Attempts - Failed\,Transactions)}{Total\,Transaction\,Attemtps}\,x100$$

Service Credit:
The following Service Levels and Service Credits are applicable to Azure AI Services APIs (except Azure OpenAI):

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Service Level Exceptions: A separate SLA applies to Azure OpenAI Service.

Table of Contents / Definitions

Azure Communication Gateway

Additional Definitions
"Assigned Telephone Number" is a telephone number that meets all the following criteria:
- It has been provisioned within the Operator Connect or Teams Phone Mobile environments.
- The telephone number is configured for connectivity through the Azure Communications Gateway.
- The telephone number's status is "assigned" in the Operator Connect or Teams Phone Mobile environments. This includes (but is not limited to) assignment to users, conferencing bridges, voice Applications and third-party applications.

"Downtime" is any period of time in an Applicable Period for a given Microsoft Azure Subscription when Assigned Telephone Numbers are unable to initiate or receive voice calls through the Azure Communications Gateway.

"Downtime Number Minutes" is the sum of all Downtime, multiplied by the number of Assigned Telephone Numbers unable to initiate or receive calls through the Azure Communications Gateway for the given Downtime.
"Maximum Available Number Minutes" is the total number of minutes in an Applicable Period that the Azure Communications Gateway has been successfully deployed (i.e. provisioning status is marked as complete) multiplied by the maximum number of assigned Telephone Numbers at any time within that Applicable Period.
"Uptime Percentage" The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Number Minutes-Downtime Number Minutes}}{\text{Maximum Available Number Minutes}} x100$$

This SLA does not apply to outages caused by any failure of third-party software, equipment, or services that are not controlled by Microsoft, or Microsoft software that is not being run as part of this Service.

The following Service Levels and Service Credits are applicable to Customer's use of Azure Communications Gateway:

Uptime Percentage	Service Credit
< 99.99%	25%
< 98%	50%
<95	100%

Table of Contents / Definitions

Azure Communication Services

Additional Definitions
"**Downtime**" is the total number of Maximum Available Minutes during an Applicable Period where Azure Communication Services is unavailable. A minute is considered unavailable if all requests within this minute results in 5xx errors.
Maximum Available Minutes is the total number of minutes that Azure Communication Services has been deployed by a Customer in a Microsoft Azure subscription during an Applicable Period.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Additional Terms: Service credit will be applied to the individual service that was unavailable. For example, if Customer is using SMS and Chat services, and the SMS service does not meet SLA, Customer would receive a credit for the SMS usage, not the chat usage.

The available minutes are based only on services that are in the control of Azure Communication Services; this excludes third party services such as telecommunications providers and carriers.

Table of Contents / Definitions

Azure Confidential Ledger

Uptime Calculation and Service Levels for Azure Confidential Ledger
"**Deployment Minutes**" is the total number of minutes that a given managed confidential ledger has been deployed in Microsoft Azure during an Applicable Period.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all managed confidential ledgers deployed by Customer in a given Microsoft Azure subscription during an Applicable Period.
"**Excluded Transactions**" are transactions for creating, updating, or deleting managed confidential ledgers.
"**Downtime**" is the total accumulated minutes, across all managed confidential ledgers deployed by Customer in a given Microsoft Azure subscription, during which the managed confidential ledgers were unavailable. A minute is considered unavailable for a given confidential ledger if all continuous attempts to perform transactions, other than Excluded Transactions, on the confidential ledger throughout the minute either return an Error Code or do not result in a Success Code within 5 seconds from Microsoft's receipt of the request.

"**Uptime Percentage**" for the Azure Confidential Ledger service is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for a given Microsoft Azure subscription. Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of Azure Confidential Ledger:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure Container Apps

Additional Definitions
"**App**" is a microservice or application deployed by a customer of the Azure Container Apps service.
"**Deployment Minutes**" means the total number of minutes that an App is expected to be active for an Applicable Period. The time that an App is expected to be active for an Applicable Period is based on scale rules set by a customer.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes for a given App deployed by Customer in a given Microsoft Azure subscription during an Applicable Period.
Uptime Calculation and Service Levels for Azure Container Apps
"**Downtime**" is the total accumulated minutes, across all Apps deployed by a Customer in a given Microsoft Azure subscription, during which one or more of the Apps is unavailable. A minute is considered unavailable for a given App when there is no connectivity between the App and Microsoft's Internet gateway.
Uptime Percentage: The "Uptime Percentage" is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Azure Container Apps service:

Uptime Percentage	Service Credit
< 99.95%	10%
< 99%	25%
< 95%	100%

Table of Contents / Definitions

Azure Container Instances

Additional Definitions:
"**Connectivity**" is bi-directional network traffic between the Container Group and other IP addresses using TCP or UDP network protocols in which the Container Group is configured for allowed traffic.
"**Container Group**" is a collection of co-located containers that shares the same lifecycle and networking resources.
Uptime Calculation and Service Levels for Container Group:
"**Maximum Available Minutes**" is the total number of minutes that a given Container Group has been deployed by Customer in a Microsoft Azure subscription during an Applicable Period. Maximum Available Minutes is measured from Customer action that results in starting a given Container Group to the time Customer action that results in stopping or deleting a given Container Group.
"**Downtime**" is the total number of minutes within Maximum Available Minutes that have no Connectivity.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credit are applicable to Customer's use of Container Group.

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure Container Registry

Additional Definitions:
"**Managed Registry**" is any instance of Basic, Standard or Premium Container Registry.
"**Registry Endpoint**" is the host name from which a given Managed Registry is accessed by clients to perform Container Registry related operations.
"**Registry Transactions**" is the set of transaction requests sent from the client to the Registry Endpoint.
Uptime Calculation and Service Levels for Managed Container Registry
"**Maximum Available Minutes**" is the total number of minutes that a given Managed Container Registry has been deployed by Customer in a Microsoft subscription during an Applicable Period.

"**Downtime**" is the total number of minutes within Maximum Available Minutes during which Managed Registry is unavailable. A minute is considered unavailable if all continuous attempts to send Registry Transactions receive an Error Code or do not respond within the Maximum Processing Time outlined in the table below.

Transaction Types	Maximum Processing Time
List (Repository, Manifests, Tags)	8 Minutes
Others	1 Minute

"**Uptime Percentage**" for Managed Container Registry is calculated using the following formula:

$$Monthly\,Uptime\% = \frac{(Maximum\,Available\,Minutes - Downtime)}{Maximum\,Available\,Minutes} x\,100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Content Delivery Network (CDN)

Uptime Calculation and Service Levels for CDN Service
Microsoft will review data from any commercially reasonable independent measurement system used by Customer.

Customer must select a set of agents from the measurement system's list of standard agents that are generally available and represent at least five geographically diverse locations in major worldwide metropolitan areas (excluding PR of China).

Measurement System tests (frequency of at least one test per hour per agent) will be configured to perform one HTTP GET operation according to the model below:

1. A test file will be placed on Customers' origin (e.g., Azure Storage account).
2. The GET operation will retrieve the file through the CDN Service, by requesting the object from the appropriate Microsoft Azure domain name hostname.
3. The test file will meet the following criteria:
   i. The test object will allow caching by including explicit "Cache-control: public" headers, or lack of "Cache-Control: private" header.
   ii. The test object will be a file at least 50KB in size and no larger than 1MB.
   iii. Raw data will be trimmed to eliminate any measurements that came from an agent experiencing technical problems during the measurement period.

"Uptime Percentage" is the percentage of HTTP transactions in which the CDN responds to client requests and delivers the requested content without error. Uptime Percentage of the CDN Service is calculated as the number of times the object was delivered successfully divided by the total number of requests (after removing erroneous data).
The following Service Levels and Service Credits are applicable to Customer's use of the CDN Service

Uptime Percentage	Service Credit
< 99.9%	10%
< 99.5%	25%

Table of Contents / Definitions

Azure Cosmos DB

SLA details enumerated for Azure Cosmos DB service include the following database APIs with distinct definitions and details for API for PostgreSQL from the remaining database APIs:

- Azure Cosmos DB for PostgreSQL
- Azure Cosmos DB for NoSQL
- Azure Cosmos DB for MongoDB RU
- Azure Cosmos DB for Apache Cassandra
- Azure Cosmos DB for Apache Gremlin
- Azure Cosmos DB for Table

Microsoft Azure Cosmos DB for PostgreSQL
"Server" is any given Azure Cosmos DB for PostgreSQL server.

"High Availability Cluster" means a set of High Availability Nodes.
"High Availability Node" means a Node within a cluster, with high availability enabled.
"Coordinator Node" is a Node that is assigned the role of Cluster Coordinator.
"Worker Node" is a Node that is assigned the role of Worker.
"Node" or **"Nodes"** is an Azure Cosmos DB for PostgreSQL Coordinator or Worker node.

Uptime Calculation and Service Levels for Microsoft Azure Cosmos DB for PostgreSQL – High Availability Node

"Maximum Available Minutes" is the total number of minutes for a given High Availability Node deployed by Customer in a Microsoft Azure subscription during an Applicable Period.
"Downtime" is the total number of minutes within Maximum Available Minutes during which a Node is unavailable. A minute is considered unavailable if all continuous attempts by Customer to establish a connection to the Node returned an Error Code or did not respond within the minute. A minute is also considered unavailable for a Worker Node if its Coordinator Node was unavailable within the minute.
"Uptime Percentage" for the Azure Cosmos DB for PostgreSQL High Availability Node is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes.
The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credit are applicable to Customer's use of the Microsoft Azure Cosmos DB for PostgreSQL High Availability Node.

Uptime Percentage	Service Credit
< 99.99%	10%
< 99%	25%

Microsoft Azure Cosmos DB for NoSQL, Microsoft Azure Cosmos DB for MongoDB (RU), Microsoft Azure Cosmos DB for Apache Cassandra, Microsoft Azure Cosmos DB for Apache Gremlin, Microsoft Azure Cosmos DB for Table

Additional Definitions:
"**Container**" is a container of data items, and a unit of scale for transactions and queries.
"**Consumed RUs**" is the sum of the Request Units consumed by all the requests which are processed by the Azure Cosmos DB Container in a given second.
"**Database Account**" is the top-level resource of the Azure Cosmos DB resource model. An Azure Cosmos DB Database Account contains one or more databases.
"**Failed Requests**" are requests within Total Requests that either return an Error Code or fail to return a Success Code within the maximum upper bounds documented in the table below.
"**Failed Read Requests**" are requests within Total Read Requests that either return an Error Code or fail to return a Success Code within the maximum upper bounds documented in the table below.

Operation	Maximum Upper Bound on Processing Latency
Resource Operations	5 Seconds
Media Operations	60 Seconds

"**Provisioned RUs**" is the total provisioned Request Units for a given Azure Cosmos DB Container for a given second.
"Provisioned Throughput Resources" are Azure Cosmos DB Containers configured in provisioned throughput mode where the number of Provisioned RUs is billed.
"**Rate Limited Requests**" are requests that return a 429 status code from the Azure Cosmos DB Container, indicating that Consumed RUs have exceeded the Provisioned RUs for a partition in the Container for a given second.
"**Request Unit (RU)**" is a measure of throughput in Azure Cosmos DB.
"**Resource**" is a set of URI addressable entities associated with a Database Account.
"Serverless Resources" are Azure Cosmos DB Containers configured in serverless mode where the number of Consumed RUs is billed.
"**Successful Requests**" are Total Requests minus Failed Requests.
"**Total Read Requests**" is the set of all the read requests, including Rate Limited Requests and all the Failed Read Requests, issued against Resources within a one-hour interval within a given Azure subscription during an Applicable Period.
"**Total Requests**" is the set of all requests, including Rate Limited Requests and all Failed Requests, issued against Resources within a one-hour interval within a given Azure subscription during an Applicable Period.

Availability SLA

"**Read Error Rate**" is the total number of Failed Read Requests divided by Total Read Requests, across all Resources in a given Azure subscription, during a given one-hour interval. If the Total Read Requests in a given one-hour interval is zero, the Read Error Rate for that interval is 0%.
"**Error Rate**" is the total number of Failed Requests divided by Total Requests, across all Resources in a given Azure subscription, during a given one-hour interval. If the Total Requests in a given one-hour interval is zero, the Error Rate for that interval is 0%.
"**Average Error Rate**" for an Applicable Period is the sum of Error Rates for each hour in the Applicable Period divided by the total number of hours in the Applicable Period.
"**Average Read Error Rate**" for an Applicable Period is the sum of Read Error Rates for each hour in the Applicable Period divided by the total number of hours in the Applicable Period.
"Availability Percentage single region" for the Azure Cosmos DB Service deployed via Database Accounts scoped to a single Azure region configured with any of the five Consistency Levels is calculated by subtracting from 100% the Average Error Rate for a given Microsoft Azure subscription in an Applicable Period.
The Availability Percentage is represented by the following formula:

100% - Average Error Rate

Service Credit for Provisioned Throughput Resources:

Availability Percentage	Service Credit
< 99.99%	10%
< 99%	25%

"Availability Percentage, single-region with availability zones (SR-AZ)" for the Azure Cosmos DB Service deployed via Database Accounts scoped to a single Azure region configured with availability zones and any of the five Consistency Levels is calculated by subtracting from 100% the Average Error Rate for a given Microsoft Azure subscription in an Applicable Period.
Availability Percentage is represented by the following formula:

100% - Average Error Rate

Service Credit for Provisioned Throughput Resources:

Availability Percentage (SR-AZ)	Service Credit
< 99.995%	10%
< 99%	25%

"Read Availability Percentage, multiple regions" for the Azure Cosmos DB Service deployed via Database Account configured to span two or more regions is calculated by subtracting from 100% the Average Read Error Rate for a given Microsoft Azure subscription in an Applicable Period.
Read Availability Percentage is represented by the following formula:

100% - Average Read Error Rate

Service Credit for Provisioned Throughput Resources:

Read Availability Percentage	Service Credit
< 99.999%	10%
< 99%	25%

"Multiple Write Locations Availability Percentage" for the Azure Cosmos DB Service deployed via Database Accounts configured to span multiple Azure regions with multiple writable locations is calculated by subtracting from 100% the Average Error Rate for a given Microsoft Azure subscription in an Applicable Period.
Availability Percentage is represented by the following formula:

Uptime % = 100% - Average Error Rate

Service Credit for Provisioned Throughput Resources:

Multiple Write Locations Availability Percentage	Service Credit
< 99.999%	10%
< 99%	25%

Throughput SLA

"**Throughput Failed Requests**" are Rate-Limited Requests resulting in an Error Code, before Consumed RUs have exceeded the Provisioned RUs for a partition in the Container for a given second.
"**Error Rate**" is the total number of Throughput Failed Requests divided by Total Requests, across all Resources in a given Azure subscription, during a given one-hour interval. If the Total Requests in a given one-hour interval is zero, the Error Rate for that interval is 0%.

"**Average Error Rate**" for an Applicable Period is the sum of Error Rates for each hour in the Applicable Period divided by the total number of hours in the Applicable Period.
"**Throughput Percentage**" for the Azure Cosmos DB Service is calculated by subtracting from 100% the Average Error Rate for a given Microsoft Azure subscription in an Applicable Period.
Throughput Percentage is represented by the following formula:

100% - Average Error Rate

Service Credit for Provisioned Throughput Resources:

Throughput Percentage	Service Credit
< 99.99%	10%
< 99%	25%

Consistency SLA

"**K**" is the number of versions of a given data item for which the reads lag behind the writes.
"**T**" is a given time interval.
"**Consistency Level**" is the setting for a particular read request that supports consistency guarantees. The following table captures the guarantees associated with the Consistency Levels. Note that Session, Bounded Staleness, Consistent Prefix and Eventual Consistency Levels are all referred to as "relaxed".

Consistency Level	Consistency Guarantees
Strong	Linearizability
Session	Read Your Own Write (within write region) Monotonic Read Consistent Prefix
Bounded Staleness	Read Your Own Write (within write region) Monotonic Read (within a region) Consistent Prefix Staleness Bound < K,T
Consistent Prefix	Consistent Prefix
Eventual	Eventual

"**Consistency Violation Rate**" is Successful Requests that could not be delivered when performing the consistency guarantees specified for the chosen Consistency Level divided by Total Requests, across all Resources in a given Azure subscription, during a given one-hour interval. If the Total Requests in a given one-hour interval is zero, the Consistency Violation Rate for that interval is 0%.
"**Average Consistency Violation Rate**" for an Applicable Period is the sum of Consistency Violation Rates for each hour in the Applicable Period divided by the total number of hours in the billing month.
"**Consistency Attainment Percentage**" for the Azure Cosmos DB Service is calculated by subtracting from 100% the Average Consistency Violation Rate for a given Microsoft Azure subscription in an Applicable Period.
Consistency Percentage: For the Azure Cosmos DB Service is calculated by subtracting from 100% the Average Consistency Violation Rate for a given Microsoft Azure subscription in an Applicable Period.
The Consistency Percentage is represented by the following formula:

100% - Average Consistency Violation Rate

Service Credit:

Consistency Attainment Percentage	Service Credit
< 99.99%	10%
< 99%	25%

Latency SLA

"**Application**" is a Azure Cosmos DB application deployed within a local Azure region with accelerated networking enabled and using the Azure Cosmos DB client SDK configured with TCP direct connectivity for a given Microsoft Azure subscription in an Applicable Period.
"**N**" is the number of Successful Requests for a given Application performing either a data item read or data item write operations with a payload size less than or equal to 1 KB in a given hour.
"**S**" is the latency-sorted set of Successful Request response times in ascending order for a given Application performing data item read or data item write operations with a payload size less than or equal to 1 KB in a given hour.
"**Ordinal Rank**" is the 99th percentile using the nearest rank method represented by the following formula:

$$Ordinal\,Rank = \frac{99}{100} x N$$

"**P99 Latency**" is the value at the Ordinal Rank of S.
"**Excessive Latency Hours**" is the total number of one-hour intervals during which Successful Requests submitted by an Application resulted in a P99 Latency greater than or equal to 10ms for data item read or 10ms for data item write operations. If the number of Successful Requests in a given one-hour interval is zero, the Excessive Latency Hours for that interval is 0.
"**Average Excessive Latency Rate**" for an Applicable Period is the sum of Excessive Latency Hours divided by the total number of hours in the Applicable Period.
"**P99 Latency Attainment Percentage**" for a given Azure Cosmos DB Application deployed via Database Accounts scoped to a single Azure region configured with any of the five Consistency Levels or Database Accounts spanning multiple regions, configured with any of the four relaxed Consistency Levels is calculated by subtracting from 100% the Average Excessive Latency Rate for a given Microsoft Azure subscription in an Applicable Period. P99 Latency Attainment Percentage is represented by the following formula:

$$100\% - \text{Average Excessive Latency Rate}$$

Service Credit for Provisioned Throughput Resources:

P99 Latency Attainment Percentage	Service Credit
< 99.99%	10%
< 99%	25%

Table of Contents / Definitions

Data Catalog

Additional Definitions:
"**Deployment Minutes**" is the total number of minutes for which a Data Catalog has been purchased during an Applicable Period.
"**Entries**" means any catalog object registration in the Data Catalog (such as a table, view, measure, cluster or report).
"**Maximum Available Minutes**" is the sum of all Deployment Minutes for the Data Catalog associated with a given Microsoft Azure subscription during an Applicable Period.
Downtime: is the total accumulated Deployment minutes, during which the Data Catalog is unavailable. A minute is considered unavailable for a given Data Catalog if all attempts by administrators to add or remove users to the Data Catalog or all attempts by users to execute API calls to the Data Catalog for registering, searching, or deleting Entries either result in an Error Code or do not return a response within five minutes.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure Data Explorer (Kusto)

Additional Definitions
"**Cluster**" means Azure Data Explorer (ADX) operated cluster.
Uptime Calculation and Service Levels for Azure Data Explorer
"**Maximum Available Minutes**" is the total number of minutes for a given Cluster deployed by Customer in a Microsoft Azure subscription during an Applicable Period.
"**Downtime**" is the total number of minutes within Maximum Available Minutes during which a Cluster is unavailable. A minute is considered unavailable for a given Cluster if all continuous attempts within the minute to establish a connection to the Cluster returned an Error Code.
"**Uptime Percentage**" for the Azure Data Explorer is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes.
Uptime Percentage is represented by the following formula:

$$\frac{Maximum\,Available\,Minutes - Downtime}{Maximum\,Available\,Minutes} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Service:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure Data Factory

Additional Definitions:
"**Resources**" means integration runtimes (including Azure, SSIS and self-hosted Integration Runtimes), triggers, pipelines, data sets, and linked services created within a Data Factory.
"**Activity Run**" means the execution or attempted execution of an activity
Uptime Calculation for Data Factory API Calls
Additional Definitions:
"**Total Requests**" is the set of all requests, other than Excluded Requests, to perform operations against Resources during an Applicable Period for a given Microsoft Azure subscription.
"**Excluded Requests**" is the set of requests that result in an HTTP 4xx status code, other than an HTTP 408 status code.
"**Failed Requests**" is the set of all requests within Total Requests that either return an Error Code or an HTTP 408 status code or otherwise fail to return a Success Code within two minutes.
"Uptime Percentage" for the API calls made to the Data Factory Services is calculated as Total Requests less Failed Requests divided by Total Requests in an Applicable Period for a given Microsoft Azure subscription.
Uptime Percentage is represented by the following formula:

$$Monthly\,Uptime\% = \frac{(Total\,Requests - Failed\,Requests)}{Total\,Requests}$$

The following Service Credits are applicable to Customer's use of API calls within the Data Factory Service

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Uptime Calculation for Data Factory Activity Runs
Additional Definitions:
"**Total Activity Runs**" is the total number of Activity Runs attempted during in an Applicable Period for a given Microsoft Azure Subscription.
"**Delayed Activity Runs**" is the total number of attempted Activity Runs in which an activity fails to begin executing within four (4) minutes after the time at which it is scheduled for execution and all dependencies that are prerequisite to execution have been satisfied.
"Uptime Percentage" for the Data Factory Service is calculated as Total Activity Runs less Delayed Activity Runs divided by Total Activity Runs in an Applicable Period for a given Microsoft Azure subscription.
The Uptime Percentage is calculated using the following formula:

$$\frac{Total\,Activity\,Runs - Delayed\,Activty\,Runs}{Total\,Activity\,Runs}\,x100$$

The following Service Levels and Service Credits are applicable to Customer's Activity Runs within the Data Factory Service:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Data Lake Analytics

Additional Definitions:
"**Total Operations**" is the total number of authenticated operations attempted within a one-hour interval across all Data Lake Analytics accounts in a given Azure subscription during an Applicable Period.
"**Failed Operations**" is the set of all operations within Total Operations that either return an Error Code or fail to return a Success Code within 5 minutes for account creation and deletion and 25 seconds for all other operations with an additional 2 seconds per MB for operations with payload.
"**Error Rate**" is the total number of Failed Operations divided by Total Operations during a given one-hour interval. If the Total Operations in a one-hour interval is zero, the Error Rate for that interval is 0%.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

100% - Average Error Rate

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Data Lake Storage Gen1

Additional Definitions:
"**Total Operations**" is the total number of authenticated operations attempted within a one-hour interval across all Data Lake Store accounts in a given Azure subscription during an Applicable Period.
"**Failed Operations**" is the set of all operations within Total Operations that either return an Error Code or fail to return a Success Code within 5 minutes for account creation and deletion, 2 seconds per file for operations on multiple files, 2 seconds per MB for data transfer operations, and 2 seconds for all other operations.
"**Error Rate**" is the total number of Failed Operations divided by Total Operations during a given one-hour interval. If the Total Operations in a one-hour interval is zero, the Error Rate for that interval is 0%.
"**Average Error Rate**" for an Applicable Period is the sum of Error Rates for each hour in the Applicable Period divided by the total number of hours in the Applicable Period.
"Uptime Percentage" is calculated by subtracting from 100% the Average Error Rate for a given Microsoft Azure subscription in an Applicable Period.
The Uptime Percentage is calculated using the following formula:

100% - Average Error Rate

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure Database for MariaDB

Additional Definitions
"**Server**" is any given Azure Database for MariaDB server.
Uptime Calculation and Service Levels for Microsoft Azure Database for MariaDB
"**Maximum Available Minutes**" is the total number of minutes for a given Server deployed by Customer in a Microsoft Azure subscription during an Applicable Period.
"**Downtime**" is the total number of minutes within Maximum Available Minutes during which a Server is unavailable. A minute is considered unavailable if all continuous attempts by Customer to establish a connection to the Server returned an Error Code.
"**Uptime Percentage**" for the Azure Database for MariaDB is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes.
Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Microsoft Azure Database for MariaDB:

Uptime Percentage	Service Credit
< 99.99%	10%
< 99%	25%
< 95%	100%

Table of Contents / Definitions

Azure Database for MySQL

Microsoft Azure Database for MySQL – Single Server
Additional Definitions:
"**Server**" is any given Azure Database for MySQL server – Single Server.
Uptime Calculation and Service Levels for Microsoft Azure Database for MySQL – Single Server
"**Maximum Available Minutes**" is the total number of minutes for a given Server deployed by Customer in a Microsoft Azure subscription during an Applicable Period.
"**Downtime**" is the total number of minutes within Maximum Available Minutes during which a Server is unavailable. A minute is considered unavailable if all continuous attempts by Customer to establish a connection to the Server returned an Error Code.
"Uptime Percentage" for the Azure Database for MySQL is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes.
The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of Azure Database for MySQL – Single Server:

Uptime Percentage	Service Credit
< 99.99%	10%
< 99%	25%
< 95%	100%

Microsoft Azure Database for MySQL – Flexible Server
Additional Definitions:
"**Server**" is any given Azure Database for MySQL server – Flexible server.
"High Availability" in the context of Flexible server means a set of High Availability servers (Primary and standby) deployed in zone redundancy or same-zone redundancy.
Uptime Calculation and Service Levels for Microsoft Azure Database for MySQL – Flexible Server
"Maximum Available Minutes" is the total number of minutes for a given Server deployed by Customer in a Microsoft Azure subscription during an Applicable Period.
"Downtime" is the total number of minutes within Maximum Available Minutes during which a Server is unavailable. A minute is considered unavailable if all continuous attempts by Customer to establish a connection to the Server were unsuccessful.
"Uptime Percentage" for the Azure Database for MySQL – Flexible Server is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes.
Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of Azure Database for MySQL – Flexible Server configured in zone-redundant High Availability mode:

Uptime Percentage	Service Credit
Less than 99.99% and greater than or equal to 99.00%	10%
Less than 99.00% and greater than or equal to 95.00%	25%
< 95%	100%

The following Service Levels and Service Credits are applicable to Customer's use of Azure Database for MySQL – Flexible Server configured in same-zone High Availability mode:

Uptime Percentage	Service Credit
Less than 99.95% and greater than or equal to 99.00%	10%
< 99%	25%

The following Service Levels and Service Credits are applicable to Customer's use of Azure Database for MySQL – Flexible Server that is not configured High Availability mode:

Uptime Percentage	Service Credit
Less than 99.9% and greater than or equal to 99.00%	10%
< 99%	25%

Azure Database for PostgreSQL

Azure Database for PostgreSQL - Single Server
Additional Definitions:
"Server" is any given Azure Database for PostgreSQL server - Single Server.
"High Availability Cluster" means a set of High Availability Nodes.
"High Availability Node" means a Node within a server group, with high availability enabled.
"Coordinator Node" is a Node that is assigned the role of Cluster Coordinator.
"Worker Node" is a Node that is assigned the role of Worker.
Uptime Calculation and Service Levels for Microsoft Azure Database for PostgreSQL - Single Server
"**Maximum Available Minutes**" is the total number of minutes for a given Server deployed by Customer in a Microsoft Azure subscription during an Applicable Period.
"**Downtime**" is the total number of minutes within Maximum Available Minutes during which a Server is unavailable. A minute is considered unavailable if all continuous attempts by Customer to establish a connection to the Server returned an Error Code or do not respond within a minute.
"Uptime Percentage" for the Azure Database for PostgreSQL is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes.
The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of Azure Database for PostgreSQL – Single Server:

Uptime Percentage	Service Credit
< 99.99%	10%
< 99%	25%
< 95%	100%

Microsoft Azure Database for PostgreSQL – Flexible Server
Additional Definitions:
"Server" is any given Azure Database for PostgreSQL server Flexible server.
"High Availability" in the context of Flexible server means a set of High Availability servers (Primary and standby) deployed in a zone-redundant configuration or same-zone redundancy.
Uptime Calculation and Service Levels for Microsoft Azure Database for PostgreSQL– Flexible Server
"Maximum Available Minutes" is the total number of minutes for a given Server deployed by Customer in a Microsoft Azure subscription during an Applicable Period.
"Downtime" is the total number of minutes within Maximum Available Minutes during which a Server is unavailable. A minute is considered unavailable if all continuous attempts by Customer to establish a connection to the Server were unsuccessful.
"Uptime Percentage" for the Azure Database for PostgreSQL – Flexible Server is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes.
Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of Azure Database for PostgreSQL – Flexible Server configured in a zone-redundant High Availability mode:

Uptime Percentage	Service Credit
Less than 99.99% and greater than or equal to 99.00%	10%
Less than 99.00% and greater than or equal to 95.00%	25%
< 95.00%	100%

The following Service Levels and Service Credits are applicable to Customer's use of Azure Database for PostgreSQL – Flexible Server configured in a same-zone High Availability mode:

Uptime Percentage	Service Credit
Less than 99.95% and greater than or equal to 99.00%	10%
< 99.00%	25%

The following Service Levels and Service Credits are applicable to Customer's use of Azure Database for PostgreSQL – Flexible Server that is not configured in High Availability mode:

Uptime Percentage	Service Credit
Less than 99.9% and greater than or equal to 99.00%	10%
< 99.00%	25%

Table of Contents / Definitions

Azure Databricks

Additional Definitions
"**Azure Databricks Gateway**" is a set of compute resources that proxy UI and API requests between Customer and Azure Databricks.
Uptime Calculation and Service Levels for Azure Databricks
"**Maximum Available Minutes**" is the total number of minutes across all Azure Databricks workspaces deployed by Customer in a given Microsoft Azure subscription in an Applicable Period.
"**Downtime**" is the total accumulated minutes of unavailability across all Azure Databricks workspaces deployed in a given Microsoft Azure subscription. A minute is considered unavailable for a given Azure Databricks workspace if all continual attempts within the minute to establish a connection to the Azure Databricks Gateway for the applicable workspace fail.
"**Uptime Percentage**" for the Azure Databricks Service is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes multiplied by 100. Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credit are applicable to Customer's use of Azure Databricks

Uptime Percentage	Service Credit
< 99.95%	10%
< 99%	25%

Table of Contents / Definitions

Microsoft Azure Data Manager for Energy

Additional definitions:
"Client" is the end-user facing portion of the Azure Data Manager for Energy resource.
"**Total API Requests**" refers to the total number of authenticated API requests made by the client to any of the API endpoints of their Azure Data Manager for Energy resource during an Applicable Period for a given Microsoft Azure subscription.
"**Failed API Requests**" is the set of all API requests within Total API requests that result in an Error Code.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{Total\,API\,Requests - Failed\,API\,Requests}{Total\,API\,Requests} X100$$

Service credit

Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Service level exceptions: The Service Levels and Service Credits are applicable to your use of the Standard Tier of the Azure Data Manager for Energy. The Developer Tier of the Microsoft Azure Data Manager for Energy is not subject to this SLA.

Table of Contents / Definitions

Azure DDoS Protection

Additional Definitions:
"**Maximum Available Minutes**" is the total number of minutes DDoS Protection Service is enabled for a given Microsoft Azure subscription during an Applicable Period.

"**Downtime**" is the total number of minutes within Maximum Available Minutes where protected Azure resources were not available. A minute is considered unavailable when DDoS Protection did not mitigate an attack which directly resulted in underlying Azure resources not meeting respective SLA.
"Uptime Percentage" is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes multiplied by 100.
The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Levels and Service Credits are applicable to Customer's use of Azure DDoS Protection:

Uptime Percentage	Service Credit
< 99.99%	10%
< 99.95%	25%

Table of Contents / Definitions

Azure Defender

Additional Definitions:
"**Protected Node**" is a Microsoft Azure resource, counted as a node for billing purposes that is configured for the Azure Defender.
"**Security Monitoring**" is the assessment of a Protected Node resulting in potential findings such as security health status, recommendations, and security alerts, exposed in Azure Defender.
"**Maximum Available Minutes**" is the total number of minutes during an Applicable Period that a given Protected Node has been deployed and configured for Security Monitoring.
"**Downtime**" is the total accumulated minutes during an Applicable Period for which Security Monitoring information of a given Protected Node is unavailable. A minute is considered unavailable for a given Protected Node if all continuous attempts to retrieve Security Monitoring information throughout the minute result in either an Error Code or do not return a Success Code within two minutes.
"Uptime Percentage" for Azure Defender of a given Protected Node in a given Applicable Period is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes.
The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of each Protected Node:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Defender External Attack Surface Management

Additional Definitions
"**Maximum Available Minutes**" is the total number of minutes that a given Defender EASM resource has been deployed by Customer in a Microsoft Azure subscription during an Applicable Period.
Defender External Attack Surface Management
Additional Definitions
"**Maximum Available Minutes**" is the total number of minutes that a given Defender EASM resource has been deployed by Customer in a Microsoft Azure subscription during an Applicable Period.
"**Downtime**" is the total number of minutes within Maximum Available Minutes that data in a Defender EASM resource are unavailable. A minute is considered unavailable for a given Defender EASM resource during which no HTTP operations resulted in a Success Code.
" **Query Availability Percentage**" for a given Defender EASM resource calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes multiplied by 100.
Query Availability Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of Defender External Attack Surface Management

Query Availability Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure Dev Ops

Additional Definitions:
"**Azure Pipelines**" is a feature that allows customers to build and deploy their applications in Azure DevOps Services.
"**User-Based Extensions**" means the set of Azure DevOps Services extensions published by Microsoft which are sold on a per-user basis via the Azure DevOps Marketplace.
"**Azure DevOps Services Users**" refers to the set of features and capabilities available to a user within an Azure DevOps Services account in a Customer subscription. The features and capabilities available are described on the Azure DevOps website.
Uptime Calculation and Service Levels for Azure DevOps Services Users and User-Based Extensions
"**Deployment Minutes**" is the total number of minutes for which a User or User-Based Extension has been purchased during an Applicable Period.
"**Maximum Available Minutes**" is the of all Deployment Minutes across all Users and User-Based Extensions for a given Microsoft Azure subscription during an Applicable Period.
"**Downtime**" is the total Deployment Minutes, across all Users and User-Based Extensions for a given Microsoft Azure subscription, during which the Service is unavailable. A minute is considered unavailable for a given User or User-Based Extension if all continuous HTTP requests to perform operations, other than operations pertaining to the Azure Pipelines Service, throughout the minute either result in an Error Code or do not return a response.
Uptime Percentage: for Azure DevOps Services Users and User-Based Extensions is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for a given Microsoft Azure subscription.
Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

In the event Azure DevOps Services are unavailable, Service Credits are applicable to Azure DevOps Services Users and User-Based Extensions. The following Service Levels and Service Credits are applicable to Customer's use of the Azure Pipelines Service.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Uptime Calculation and Service Levels for Azure Pipelines
"**Maximum Available Minutes**" is the total number of minutes for which the paid Azure Pipelines Service has been enabled for a given Microsoft Azure subscription during an Applicable Period.
"**Downtime**" is the total accumulated minutes for a given Microsoft Azure subscription during which the Azure Pipelines Service is unavailable. A minute is considered unavailable if all continuous HTTP requests to the Azure Pipelines Service to perform operations initiated by Customer throughout the minute either result in an Error Code or do not return a response.
"**Uptime Percentage**" for the Azure Pipelines Service is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for a given Microsoft Azure subscription.
Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Azure Pipelines Service:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Microsoft Dev Box

Additional Definitions:
"Dev Box" means the specific instance of Microsoft Dev Box.
"Minutes in the Applicable Period" means the total number of minutes in a given Applicable Period.

"Downtime" is measured in minutes, the period in which all connection attempts by a specific user to a specific Dev Box were unsuccessful, excluding any of the following types of failures:

1. Failures resulting from the Dev Box being in an inoperable state unrelated to the underlying Azure infrastructure (e.g. damaged or corrupt operating system, operating system configuration, or misconfiguration); and
2. Failure resulting from an application or other software installed on the Dev Box.

"Per-Dev Box Uptime Percentage" is calculated by subtracting from 100% the percentage of Minutes in the Applicable Period in which the Dev Box had Downtime.
The Per-Dev Box Uptime Percentage is calculated using the following formula:

$$Per-Dev\,Box\,Uptime\,\% = \frac{\left(Minutes \in the\,Applicable\,Period - Downtime\right)}{Minutes \in the\,Applicable\,Period}\,x100$$

Per-Dev Box Service Credit: The following Service Levels and Service Credits are applicable to Customer's use of Microsoft Dev Box.

Per-Dev Box Uptime Percentage	Per-Dev Box Service Credit
< 99.9%	10%
< 99%	25%
< 95%	100%

Table of Contents / Definitions

Azure Digital Twins

Additional Definitions
"**Message**" refers to any event sent by a deployed Azure Digital Twins instance to an end point service like Event Hub, Event Grid and Service Bus.
"**API Operations**" refers to read, write, update, delete and other actions performed on models and digital twins, including queries.
Uptime Calculation and Service Levels
"**Deployment Minutes**" is the total number of minutes that a given Azure Digital Twins instance has been deployed in Azure during an Applicable Period.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Azure Digital Twins instances deployed in a given Azure subscription during an Applicable Period.
"**Downtime**" is the total accumulated Deployment Minutes, across all Azure Digital Twins deployed in a given Azure subscription, during which the Azure Digital Twins instance is unavailable. A minute is considered unavailable for a given Azure Digital Twins instance if all continuous attempts to send Messages or perform API Operations on the Azure Digital Twins instance throughout the minute either return an Error Code or do not result in a Success Code within five minutes.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}}\,x100$$

The following Service Levels and Service Credits are applicable to Customer's use of Azure Digital Twins:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure DNS

Additional Definitions:
"**DNS Zone**" refers to a deployment of the Azure DNS Service containing a DNS zone and record sets.
"**Deployment Minutes**" is the total number of minutes that a given DNS Zone has been deployed in Microsoft Azure during an Applicable Period.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all DNS Zones deployed in a given Microsoft Azure subscription during an Applicable Period.

"**Valid DNS Request**" means a DNS request to an Azure DNS Service name server associated with a DNS Zone for a matching record set within the DNS Zone.
"**Downtime**" is the total accumulated Maximum Available Minutes during which the DNS Zone is unavailable. A minute is considered unavailable for a given DNS Zone if a DNS response is not received within two seconds to a valid DNS Request, provided that the valid DNS Request is made to all name servers associated with the DNS Zone and retries are continually attempted for at least 60 consecutive seconds.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{Maximum\,Available Minutes - Downtime}{Maximum\,Available Minutes} x100$$

Service Credit:

Uptime Percentage	Service Credit
<100	10%
< 99.99%	25%
< 99.5%	100%

Table of Contents / Definitions

Azure DNS Private Resolver

Additional Definitions:
"**Azure DNS Private Resolver endpoint**" refers to a deployment of an Azure DNS Private Resolver endpoint which provides name resolution for received DNS queries.
"**Deployment Minutes**" is the total number of minutes that a given endpoint has been deployed in Microsoft Azure during an Applicable Period.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all endpoints deployed in each Microsoft Azure subscription during an Applicable Period.
"**Valid DNS Request**" means a DNS request to an Azure DNS Service name server associated with a DNS Zone for a matching record set within the DNS Zone.
"**Downtime**" is the total accumulated Maximum Available Minutes during which the Azure DNS Private Resolver endpoint is unavailable. A minute is considered unavailable for a given endpoint if a DNS response is not received within two seconds to a valid DNS Request, provided that the valid DNS Request is made to all name servers associated with the endpoint and retries are continually attempted for at least 60 consecutive seconds.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{Maximum\,Available Minutes - Downtime}{Maximum\,Available Minutes} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.99%	10%
< 99.9%	25%
< 99.5%	100%

Table of Contents / Definitions

Azure DocumentDB (with MongoDB compatibility)

Additional Definitions:
"Server" is any given Azure DocumentDB server.
"High Availability Cluster" means a set of High Availability Nodes.
"High Availability Node" means a Node within a cluster, with high availability enabled.
Uptime Calculation and Service Levels for Microsoft Azure DocumentDB – High Availability Node
"Maximum Available Minutes" is the total number of minutes for a given High Availability Node deployed by Customer in a Microsoft Azure subscription during an Applicable Period.
"Downtime" is the total number of minutes within Maximum Available Minutes during which a Node is unavailable. A minute is considered unavailable if all continuous attempts by Customer to establish a connection to the Node returned an Error Code or did not respond within the minute.
"Uptime Percentage" for the Azure DocumentDB High Availability Node is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes.

The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credit are applicable to Customer's use of the Microsoft Azure DocumentDB High Availability Node configured to span two or more Azure regions.

Uptime Percentage	Service Credit
< 99.995%	10%
< 99%	25%

The following Service Levels and Service Credit are applicable to Customer's use of the Microsoft Azure DocumentDB High Availability Node scoped to a single Azure region.

Uptime Percentage	Service Credit
< 99.99%	10%
< 99%	25%

Table of Contents / Definitions

Elastic San SLA

Additional Definitions:
"Total Storage Transactions" is the set of all storage transactions attempted within a one-hour interval in a single Elastic SAN volume where the total transaction count is > 10,000. The only type of storage transaction Elastic SAN support is iSCSI operations.
"Failed Storage transactions" is the set of all read & write storage transactions within Total Storage Transactions that are not completed with a successful response due to service side failures excluding client-side errors and server throttling errors.
"Monthly Uptime Percentage": Monthly Uptime Percentage on Azure Elastic SAN on a single volume is calculated as the average of all one-hour intervals availability percentages, which is calculated using the following formula:

$$\frac{Total\,Storage\,Transactions - Failed\,Storage\,Transactions}{Total\,Storage\,Transactions} x100$$

Service Credit: Elastic SAN in LRS, ZRS (read & write requests) where the total Elastic SAN volume count > 10 and >10% of volumes within this Elastic SAN deployment do not meet SLA:

Monthly Uptime Percentage (per volume)	Service Credit (based on the Elastic SAN monthly usage)
<99.99%	5%
<99%	10%

Service Credit: Elastic SAN in LRS, ZRS (read & write requests) where the total Elastic SAN volume count > 10 and >30% of volumes within this Elastic SAN deployment do not meet SLA:

Monthly Uptime Percentage (per volume)	Service Credit
<99.99%	10%
<99%	30%

Service Credit: Elastic SAN in LRS, ZRS (read & write requests) where the total Elastic SAN volume count <=10 and any volumes within this Elastic SAN deployment do not meet SLA:

Monthly Uptime Percentage (per volume)	Service Credit
<95%	10%
<90%	30%

Table of Contents / Definitions

Event Grid

Additional Definitions:
"**Maximum Available Minutes**" is the total number of minutes that an Event Grid has been deployed by Customer in a Microsoft Azure subscription during an Applicable Period.
"**Downtime**" is the total number of minutes within Maximum Available Minutes across all Event Grids deployed by Customer in a given Microsoft Azure subscription during which Event Grid is unavailable. A minute is considered

unavailable for a given Event Grid if all requests to publish a message either return an Error Code or do not result in a Success Code within one minute.
"**Uptime Percentage**": The Uptime Percentage is calculated using the following formula:

$$\frac{Maximum\,Available\,Minutes - Downtime}{Maximum\,Available\,Minutes} x100$$

Service Credit:

Uptime Percentage	Service Credit
<99.99%	10%
<99%	25%

Table of Contents / Definitions

Geo Redundant Storage (GRS) - Priority Replication

This SLA applies to any Geographically Redundant Storage or Geographically Zone Redundant Storage accounts where Geo Priority Replication is enabled.

Additional Definitions:
"**Exceeded Last Sync Time (LST) Minute**" is each minute in a billing month the Geographically Redundant Storage Account or Geographically Zone Redundant Storage Account's Last Sync Time exceeds 15 minutes for Block Blob data.
"**Geographically Redundant Storage (GRS) Account**" is a storage account for which data is replicated within a Primary Region and then replicated asynchronously to a Secondary Region. You cannot directly read data from or write data to the Secondary Region associated with GRS Accounts.
"**Geographically Zone Redundant Storage (GZRS) Account**" is a storage account for which data is replicated across multiple facilities using Zone Redundant Storage within a Primary Region and then replicated asynchronously to a Secondary Region using Locally Redundant Storage. You cannot directly read data from or write data to the Secondary Region associated with GZRS Accounts.
"**Last Sync Time**" indicates the last time all full data synchronization from the Primary Region was replicated successfully to the Secondary Region. All writes made to the Primary Region before the Last Sync Time are available to be read from the Secondary Region.
"**Locally Redundant Storage (LRS) Account**" is a storage account for which data is replicated only within a Primary Region.
"**Maximum Available Minutes**" is the total number of minutes that a Geographically Redundant Storage Account or Geographically Zone Redundant Storage Account with Geo Priority Replication enabled has been deployed by Customer in a Microsoft Azure subscription during a billing month.
"**Primary Region**" is a geographical region in which data within a storage account is located, as selected by you when creating the storage account. You may execute write requests only against data stored within the Primary Region associated with storage accounts.
"**Read Access Geo Redundant Storage (RA-GRS) Account**" is a storage account for which data is replicated within a Primary Region and then replicated asynchronously to a Secondary Region. You can directly read data from, but cannot write data to, the Secondary Region associated with RA-GRS Accounts.
"**Read Access Geo Zone Redundant Storage (RA-GZRS) Account**" is a storage account for which data is replicated across multiple facilities using Zone Redundant Storage within a Primary Region and then replicated asynchronously to a Secondary Region using Locally Redundant Storage. You can directly read data from, but cannot write data to, the Secondary Region associated with RA-GZRS Accounts.
"**Secondary Region**" is a geographical region in which data within a GRS or GZRS Account is replicated and stored, as assigned by Microsoft Azure based on the Primary Region associated with the storage account. You cannot specify the Secondary Region associated with storage accounts.
"**Zone Redundant Storage (ZRS) Account**" is a storage account for which data is replicated across multiple facilities. These facilities may be within the same geographical region or across two geographical regions.
"Monthly **Conformance Percentage**": The Monthly Conformance Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Exceeded Last Sync Time Minute}}{\text{Maximum Available Minutes}} x100$$

Service Credit:

Conformance Percentage	Service Credit
Less than 99.00% and greater than or equal to 95.00%	10%

Conformance Percentage	Service Credit
Less than 95.00% and greater than or equal to 90.00%	25%
<90.00%	100%

Service Level Exceptions: This SLA does not apply to (i) Storage accounts that do not have Geo Priority Replication enabled, (ii) Storage accounts where append or page blob API calls were made within the last 30 days, (iii) Storage accounts with a Last Sync Time greater than 15 minutes lagged when enabling Geo Priority Replication will not be eligible until the account's Last Sync Time is less than or equal to 15 minutes lagged or (iv) during time periods where (a) Your data transfer rate exceeds 1 gigabit per second (Gbps) and the resulting backlog of writes are being replicated or (b) Your copy blob request exceeds 100 CopyBlob requests per second and the resulting backlog of writes are being replicated.

Table of Contents / Definitions

Event Hubs

Additional Definitions:
"**Message**" refers to any user-defined content sent or received through Service Bus Relays, Queues, Topics, or Notification Hubs, using any protocol supported by Service Bus.

Uptime Calculation and Service Levels for Event Hubs in the Basic and Standard tiers
"**Deployment Minutes**" is the total number of minutes that a given Event Hub has been deployed in Microsoft Azure during an Applicable Period.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Event Hubs deployed by Customer in a given Microsoft Azure subscription under the Basic or Standard Event Hubs tiers during an Applicable Period.
Downtime: The total accumulated Deployment Minutes, across all Event Hubs deployed by you in a given Microsoft Azure subscription under the Basic or Standard Event Hubs tiers, during which the Event Hub is unavailable. A minute is considered unavailable for a given Event Hub if all continuous attempts to send or receive Messages or perform other operations on the Event Hub throughout the minute either return an Error Code or do not result in a Success Code within five minutes.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Basic and Standard Event Hubs tiers:

Uptime Percentage	Service Credit
< 99.95%	10%
< 99%	25%

Uptime Calculation and Service Levels for Event Hubs in the Premium and Dedicated tiers
"**Deployment Minutes**" is the total number of minutes that a given Event Hub has been deployed in Microsoft Azure during an Applicable Period.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Event Hubs deployed by Customer in a given Microsoft Azure subscription under the Premium or Dedicated Event Hubs tiers during an Applicable Period.
"**Downtime**" is the total accumulated Deployment Minutes, across all Event Hubs deployed by Customer in a given Microsoft Azure subscription under the Premium or Dedicated Event Hubs tiers, during which the Event Hub is unavailable. A minute is considered unavailable for a given Event Hub if all continuous attempts to send or receive Messages or perform other operations on the Event Hub throughout the minute either return an Error Code or do not result in a Success Code within five minutes.
"**Uptime Percentage**" for Event Hubs is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for a given Microsoft Azure subscription.
Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Premium or Dedicated tiers:

Uptime Percentage	Service Credit
< 99.99%	10%
< 99%	25%

Table of Contents / Definitions

Azure ExpressRoute

Additional Definitions:
"**Connection**" is the Azure object that links the ExpressRoute Gateway to a Dedicated Circuit. For this SLA an ExpressRoute Gateway can have Connections to multiple Dedicated Circuits.
"**Dedicated Circuit**" means a logical representation of connectivity offered through the ExpressRoute Service between your premises and Microsoft Azure through an ExpressRoute connectivity provider or directly via an ExpressRoute Direct Port, where such connectivity does not traverse the public Internet.
"**Downtime**" is the total accumulated minutes during an Applicable Period for a given Microsoft Azure subscription during which the Dedicated Circuit is unavailable. A minute is considered unavailable for a given Dedicated Circuit if all attempts by you within the minute to establish IP-level connectivity to the Express Route Gateway associated with the Virtual Network fail for longer than thirty seconds. Traffic flowing across an alternate Dedicated Circuit in a 2+ Site Configuration would not be considered downtime for the down Dedicated Circuit in this configuration.
"**ER Gateway**" refers to a Virtual Network Gateway that facilitates connectivity between a Virtual Network and a Dedicated Circuit.
"**Metro Peering Location**" is the city name including the label "Metro" indicating that a Dedicated Circuit with this Metro Peering Location lands in the two peering locations within that city. E.g. Washington DC Metro, Sydney Metro would have links to two non-Metro peering locations in those cities.
"**Maximum Available Minutes**" is the total number of minutes that a given Dedicated Circuit is linked to one or more Virtual Networks in Microsoft Azure during an Applicable Period in a given Microsoft Azure subscription.
"**Virtual Network**" refers to a virtual private network that includes a collection of user-defined IP addresses and subnets that form a network boundary within Microsoft Azure.

There are three categories to the ExpressRoute SLA:

- "**2+ Site Configuration**" – connections landing in two or more different peering locations with non-Metro Dedicated Circuits. An example of connections to sites that would qualify for the 2+ site configurations would be Washington DC2 and Chicago, or Chicago and Chicago 2, etc. Adding a third different site qualifies under this configuration. Connections to a non-metro and metro location would be considered a "Metro site" configuration and not a 2+ Site Configuration.
- "**Metro Site Configuration**" – one (or more) connection from an ER Gateway to an ExpressRoute circuit(s) in a single Metro Peering Location.
- "**Single Site Configuration**" – one (or more) connections from an ER Gateway to an ExpressRoute circuit(s) land in one and only one (non-Metro) Peering Location.

Service Credit The following Service Levels and Service Credits are applicable to Customer's use of each Dedicated Circuit within the ExpressRoute Service depending on the site configuration. Service Credits only apply in a multi-circuit or site architecture if both circuit and/or sites are down simultaneously resulting in total loss of connectivity to the connected gateway.

At least two Dedicated Circuits in a 2+ Site Configuration to an ER Gateway:

Uptime Percentage	Service Credit
< 99.95%	10%
< 99.9%	25%

Dedicated Circuit in a Metro Site Configuration to an ER Gateway:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Dedicated Circuit in a Single Site Configuration to an ER Gateway:

Uptime Percentage	Service Credit
< 99.0%	10%
< 95%	25%

Table of Contents / Definitions

Azure ExpressRoute Traffic Collector

Additional Definitions:
"**Dedicated Circuit**" means a logical representation of connectivity offered through the ExpressRoute Service between your premises and Microsoft Azure through an ExpressRoute Direct connectivity, where such connectivity does not traverse the public Internet.
"ExpressRoute Traffic Collector" refers to a traffic collector that facilitates collection of flow logs about the IP traffic traversing over the dedicated circuit.
"**Maximum Available Minutes**" is the total number of minutes that a given ExpressRoute Traffic Controller Circuit is linked to one or more Dedicated Circuits in Microsoft Azure during an Applicable Period in a given Microsoft Azure subscription.
"**Downtime**" is the total number of minutes within Maximum Available Minutes that data in ExpressRoute Traffic Collector are unavailable during which no collected flow records is delivered for longer than 5 minutes.
"**Availability Percentage**" is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit The following Service Levels and Service Credits are applicable to Customer's use of Azure ExpressRoute Traffic Collector:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure Files Premium Tier

Additional Definitions
"File Share" is a logical storage resource in Azure Files that contains a file system and is used to store data.
"**Locally Redundant Storage (LRS)**" is a setting which indicates that data is replicated synchronously only within a Primary Region.
"**Zone Redundant Storage (ZRS)"** is a setting which indicates that data is replicated across multiple facilities. These facilities may be within the same geographical region or across two geographical regions.
"Maximum Available Minutes" is the total number of minutes a given File Share is deployed by Customer in a given Microsoft Azure subscription during an Applicable Period.
"Service Side Issue" is reported when a request fails with response type ServerOtherError or ServerBusyError or ServerTimeoutError.
"Downtime" is the total number of minutes in the Applicable Period during which all requests against the File Share have failed due to a Service Side Issue.
"Uptime Percentage": Uptime Percentage is calculated using the following formula:

$$\frac{Maximum\,Available Minutes - Downtime}{Maximum\,Available Minutes} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of File Share with Premium Tier using Zone Redundant Storage (ZRS) or Locally Redundant Storage (LRS).

Uptime Percentage	Service Credit
< 99.99%	10%
< 99%	25%

Table of Contents / Definitions

Azure Firewall

Additional Definitions:
"**Azure Firewall Service**" refers to a logical firewall instance deployed in a customer Virtual Network.
Uptime Calculation and Service Levels for the Azure Firewall Service deployed within a single Availability Zone
"**Maximum Available Minutes**" is the total accumulated minutes in an Applicable Period during which the Azure Firewall Service has been deployed in a Microsoft Azure subscription.

"**Downtime**" is the total accumulated Maximum Available Minutes in an Applicable Period for a given Azure Firewall Service during which the Azure Firewall Service is unavailable. A given minute is considered unavailable if all attempts to connect to the Azure Firewall Service throughout the minute are unsuccessful.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{Maximum\,Available\,Minutes - Downtime}{Maximum\,Available\,Minutes} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Azure Firewall Service, when deployed within a single Availability Zone:

Uptime Percentage	Service Credit
< 99.95%	10%
< 99%	25%

Uptime Calculation and Service Levels for the Azure Firewall Service deployed within two or more Availability Zones
"Maximum Available Minutes" is the total accumulated minutes in an Applicable Period during which the Azure Firewall Service has been deployed within two or more Availability Zones in the same region in a Microsoft subscription.
"Downtime" is the total accumulated Maximum Available Minutes in an Applicable Period for a given Azure Firewall Service, deployed within two or more Availability Zones, during which the Azure Firewall Service is unavailable. A given minute is considered unavailable if all attempts to connect to the Azure Firewall Service throughout the minute are unsuccessful.
"Uptime Percentage" for Azure Firewalls deployed within two or more Availability Zones is calculated using the following formula:

$$\frac{Maximum\,Available\,Minutes - Downtime}{Maximum\,Available\,Minutes} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Azure Firewall Service, when deployed within two or more Availability Zones in the same region:

Uptime Percentage	Service Credit
< 99.99%	10%
< 99%	25%

Table of Contents / Definitions

Azure Fluid Relay

Additional Definitions
"**Maximum Available Minutes**" is the total accumulated minutes in an Applicable Period during which at least one Azure Fluid Relay resource has been deployed in a Microsoft Azure subscription.
"**Downtime**" is the total accumulated Maximum Available Minutes in an Applicable Period during which at least one Azure Fluid Relay resource has been deployed, but the service calls for the Azure Fluid Relay resource are unavailable.
"**Uptime Percentage**" The Uptime Percentage is calculated using the following formula:

$$\frac{Maximum\,Available\,Minutes - Downtime}{Maximum\,Available\,Minutes} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of Azure Fluid Relay resource:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Service Level Exceptions: No SLA is provided for the Basic Tier.

Table of Contents / Definitions

Azure Front Door and Azure Front Door (classic)

Uptime Calculation and Service Levels for Azure Front Door and Azure Front Door (classic)

Microsoft will review data from any commercially reasonable independent measurement system used by Customer. Customer must select a set of agents from the measurement system's list of standard agents that are generally available and represent at least five geographically diverse locations in major worldwide metropolitan areas (excluding PR of China).

1. Measurement System tests (frequency of at least one test per 5 minutes per agent) will be configured to perform one HTTP GET operation according to the model below:
2. A test file will be placed on Customer's backend (e.g., Azure Storage account).
3. The GET operation will retrieve the file through Azure Front Door and Azure Front Door (classic), by requesting the object from the appropriate Microsoft Azure domain name hostname.
4. The test file will meet the following criteria:
   - The test object will be a file at least 50KB in size.
   - Raw data will be trimmed to eliminate any measurements that came from an agent experiencing technical problems during the measurement period.

"**Uptime Percentage**" is the percentage of HTTP transactions in which Azure Front Door and Azure Front Door (classic) responds to client requests and delivers the requested content without error. Uptime Percentage of Azure Front Door and Azure Front Door (classic) is calculated as the number of times the object was delivered successfully divided by the total number of requests (after removing erroneous data).

The following Service Levels and Service Credits are applicable to Customer's use of Azure Front Door and Azure Front Door (classic):

Uptime Percentage	Service Credit
< 99.99%	10%
< 99.9%	25%

Table of Contents / Definitions

Azure Functions

Additional Definitions

"**Function App**" is a collection of one or more functions deployed with an associated trigger.

Uptime Calculation and Service Levels for Function App on the Consumption Plan

"**Total Triggered Executions**" is the total number of all Function App executions triggered by Customer in a given Microsoft Azure subscription during an Applicable Period.

"**Unavailable Executions**" is the total number of executions within Total Triggered Executions which failed to run. An execution failed to run when the given Function App history log did not capture any output five (5) minutes after the trigger is successfully fired.

"**Uptime Percentage**" for Function Apps on the Consumption plan is calculated as Total Triggered Executions less Unavailable Executions divided by Total Triggered Executions multiplied by 100.

$$\frac{Total\ Triggered\ Executions - Unavailable\ Executions}{Total\ Triggered\ Executions} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of Function App on the Consumption plan.

Uptime Percentage	Service Credit
< 99.95%	10%
< 99%	25%
< 95%	100%

Uptime Calculation and Service Levels for Function Apps on the Flex Consumption Plan, Premium Plan, or the Dedicated App Service Plan

"**Deployment Minutes**" is the total number of minutes that a given Function App is available to be triggered during an Applicable Period. Deployment Minutes are measured based on the total time that the service is available to trigger a function execution and not based on the potential number of function executions that might be triggered during a given Applicable Period.

"**Maximum Available Minutes**" is the sum of all Deployment Minutes for a given Function App deployed by Customer in a given Microsoft Azure subscription during an Applicable Period.

"**Downtime**" is the total number of minutes within Maximum Available Minutes, during which the Function App is unavailable to be triggered. A minute is considered unavailable for a given Function App when there is no connectivity between plan on which the Function App is hosted (the Flex Consumption plan, Premium plan, or the Dedicated App Service plan) and Microsoft's Internet gateway.

"**Uptime Percentage**" for Function Apps on the Flex Consumption plan, Premium plan, or the Dedicated App Service plan is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes multiplied by 100.

$$\frac{Maximum\,Available\,Minutes - Downtime}{Maximum\,Available\,Minutes} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.95%	10%
< 99%	25%
< 95%	100%

Table of Contents / Definitions

Global Secure Access

Microsoft Entra Private Access and Microsoft Entra Internet Access
Additional Definitions
Downtime: Any period of time when Global Secure Access is unavailable to accept or process Transactions or Sessions. Downtime does not include the unavailability of the web user interface (UI), APIs, and log processing. A given minute is considered unavailable if all attempts to connect to GSA throughout the minute are unsuccessful or even if connecting, GSA is not able to process Transactions or Sessions.
Transaction: An HTTP or HTTPS request sent to or from an end user's use of the GSA.
Session: Any non-HTTP or non-HTTPS request to or from an end user's use of the GSA.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{User Minutes-Downtime}}{\text{User Minutes}} x100$$

where Downtime is measured in user minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each incident that occurs during that Applicable Period multiplied by the number of users impacted by that incident.
Service Level Exceptions:

- Performance or availability issues that result from:
  - route convergence time if using Border Gateway Protocol or Anycast;
  - scaling events and scheduled maintenance, including switchover rime during high availability events.

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
<95%	100%

Table of Contents / Definitions

HDInsight

Additional Definitions:
"**Cluster Internet Gateway**" means a set of virtual machines within an HDInsight Cluster that proxy all connectivity requests to the Cluster.
"**Deployment Minutes**" is the total number of minutes that a given HDInsight Cluster has been deployed in Microsoft Azure.
"**HDInsight Cluster**" or "**Cluster**" means a collection of virtual machines running a single instance of the HDInsight Service.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Clusters deployed by you in a given Microsoft Azure subscription during an Applicable Period.
Downtime: The total accumulated Deployment Minutes when the HDInsight Service is unavailable. A minute is considered unavailable for a given Cluster if all continual attempts within the minute to establish a connection to the Cluster Internet Gateway fail.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure Health Data Services (excluding the MedTech Service)

Additional Definitions:
"**Total Transaction Attempts**" is the total number of authenticated API requests by Customer during an Applicable Period for a given Health Data Services API (excluding MedTech service). Total Transaction Attempts do not include API requests that return an Error Code that are continuously repeated within a five-minute window after the first Error Code is received.
"**Failed Transactions**" is the set of all requests to the Health Data Services APIs (excluding MedTech service) within Total Transaction Attempts that return an Error Code. Failed Transaction Attempts do not include API requests that return an Error Code that are continuously repeated within a five-minute window after the first Error Code is received.
Uptime Calculation
Uptime Percentage: for each API service (excluding the MedTech service) is calculated as Total Transaction Attempts less Failed Transactions divided by Total Transaction Attempts. Uptime Percentage is represented by the following formula:

$$\frac{Total\ Transaction\ Attempts - Failed\ Transactions}{Total\ Transaction\ Attempts} x100$$

The following Service Levels and Service Credits are applicable to Azure Health Data Services (excluding the MedTech service):
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Health Bot

Additional Definitions
"**Azure Health Bot Premium Channel**" is a Bot Framework channel in the premium category including Webchat and Direct Line.
"**Health Bot Customer Application**" is the customer's Internet facing, conversational Health Bot application which is registered with and is configured to send and receive messages from the Azure Health Bot Service.
"**Health Bot Client**" is the end user facing portion of a Health Bot Customer Application.
"**Azure Health Bot**" is a platform for building, connecting, testing, and deploying powerful and intelligent virtual assistants.
"**Azure Health Bot Channels API Endpoint**" is a REST API endpoint that the Health Bot Client uses for HTTP communication over Health Bot Channels.
"**Total API Requests**" is the total number of HTTP requests made by the Health Bot Customer Application or the Health Bot Client to the Azure Health Bot Channels API Endpoint in during an Applicable Period.
"**Failed API Reque**sts" are the total number of requests within Total API Requests that return an Error Code or do not respond within 2 minutes.
"**Uptime Percentage**" is calculated as Total API Requests less Failed API Requests divided by Total API Requests multiplied by 100.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Total API Requests-Failed API Requests}}{\text{Total API Requests}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Microsoft Health Bot Channels.

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure Information Protection

Downtime: Any period of time when end users cannot create or consume IRM documents and email.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Azure IoT Central

Additional Definitions:
"**Deployment Minutes**" is the total number of minutes that a given IoT Central application has been deployed in given Microsoft Azure Subscription during an Applicable Period.
"**Device Identity Operations**" refers to create, read, update, and delete operations performed on the devices of an IoT Central application.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all IoT Central applications deployed in a given Microsoft Azure subscription during an Applicable Period.
"**Message**" refers to any content sent by a deployed IoT Central application to a device registered to the IoT Central application or received by the IoT Central application from a registered device.
Downtime: The total accumulated Maximum Available Minutes during which IoT Central is unavailable. A minute is considered unavailable for a given IoT Central application if all continuous attempts to send or receive Messages or perform Device Identity Operations on the IoT Central application throughout the minute either return an Error Code or do not result in a Success Code within five minutes.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure IoT Hub

Uptime Calculation and Service Levels for IoT Hub
Additional Definitions:
"**Deployment Minutes**" is the total number of minutes that a given IoT hub has been deployed in Microsoft Azure during an Applicable Period.

"**Device Identity Operations**" refers to create, read, update, and delete operations performed on the device identity registry of an IoT hub.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all IoT hubs deployed in a given Microsoft Azure subscription during an Applicable Period.
"**Message**" refers to any content sent by a deployed IoT hub to a device registered to the IoT hub or received by the IoT hub from a registered device, using any protocol supported by the Service.
Downtime: The total accumulated Deployment Minutes, across all IoT hubs deployed in a given Microsoft Azure subscription, during which the IoT hub is unavailable. A minute is considered unavailable for a given IoT hub if all continuous attempts to send or receive Messages or perform Device Identity Operations on the IoT hub throughout the minute either return an Error Code or do not result in a Success Code within five minutes.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of IoT Hub:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Uptime Calculation and Service Levels for IoT Hub Device Provisioning Service
Additional Definitions:
"**Maximum Available Minutes**" is the total number of minutes for a given Device Provisioning Service deployed by the Customer in a Microsoft Azure subscription during an Applicable Period.
"**Downtime**" is the total number of minutes within the Maximum Available Minutes during which Device Provisioning Service is unavailable. A minute is considered unavailable for a given Device Provisioning Service if all continuous attempts to register a device or perform enrollment/registration record operations on the Device Provisioning Service throughout the minute either return an Error Code or do not result in a Success Code within two minutes.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of IoT Hub Device Provisioning Service:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Key Vault

Additional Definitions:
"**Deployment Minutes**" is the total number of minutes that a given key vault has been deployed in Microsoft Azure during a billing month.
"**Excluded Transactions**" are transactions for creating, updating, or deleting key vaults, keys, or secrets.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Key Vaults deployed by you in a given Microsoft Azure subscription during a billing month.
Downtime: is the total accumulated Deployment Minutes, across all key vaults deployed by Customer in a given Microsoft Azure subscription, during which the key vault is unavailable. A minute is considered unavailable for a given key vault if all continuous attempts to perform transactions, other than Excluded Transactions, on the key vault throughout the minute either return an Error Code or do not result in a Success Code within 5 seconds from Microsoft's receipt of the request.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.99%	10%

Uptime Percentage	Service Credit
< 99%	25%

Table of Contents / Definitions

Azure Key Vault Managed HSM

Uptime Calculation and Service Levels for Managed HSM

"**Deployment Minutes**" is the total number of minutes that a given managed HSM has been deployed in Microsoft Azure during an Applicable Period.

"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all managed HSMs deployed by Customer in a given Microsoft Azure subscription during an Applicable Period.

"**Excluded Transactions**" are transactions for creating, updating, or deleting managed HSMs, keys, role assignments, role definitions, and downloading/uploading security domain.

"**Downtime**" is the total accumulated Deployment Minutes, across all managed HSMs deployed by Customer in a given Microsoft Azure subscription, during which the managed HSM is unavailable. A minute is considered unavailable for a given managed HSM if all continuous attempts to perform transactions, other than Excluded Transactions, on the managed HSM throughout the minute either return an Error Code or do not result in a Success Code within 5 seconds from Microsoft's receipt of the request.

"**Uptime Percentage**" for the Managed HSM Service is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for a given Microsoft Azure subscription.

Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Managed HSM Service:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

For the Managed HSM Service, where a managed HSM pool is replicated across Azure regions, the following Service Levels and corresponding Service Credits are applicable (Uptime Percentage is calculated based on the combined availability of the primary and replicated HSM pools):

Uptime Percentage	Service Credit
< 99.99%	10%
< 99%	25%

Table of Contents / Definitions

Azure Kubernetes Service (AKS)

Additional Definitions

"**Azure Kubernetes Service (AKS) Cluster**" A Kubernetes cluster is divided into two components:

- Control plane nodes provide the core Kubernetes services and orchestration of application workloads.
- Nodes run application workloads.

"**Kubernetes API Server**" When you create an Azure Kubernetes Service (AKS) Cluster, a control plane is automatically created and configured. The control plane includes the API Server that exposes the underlying Kubernetes API.

"**Availability Zone**" is a fault-isolated area within an Azure region, providing redundant power, cooling, and networking.

Uptime Calculation and Service Levels for AKS Clusters that use Availability Zones

"**Maximum Available Minutes**" is the total accumulated minutes of an Availability Zone enabled AKS Cluster to the time Customer has initiated an action to stop or delete the AKS Cluster during an Applicable Period.

"**Downtime**" is the total accumulated minutes that are part of Maximum Available Minutes with no connectivity to Kubernetes API Server from provisioned Availability Zone enabled AKS Cluster in the region.

"**Uptime Percentage**" for Availability Zone enabled AKS Clusters is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for a given Microsoft Azure subscription.

Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of AKS Clusters that have Availability Zones enabled in the region:

Uptime Percentage	Service Credit
< 99.95%	10%
< 99%	25%
< 95%	100%

Uptime Calculation and Service Levels for AKS Clusters that don't use Availability Zones
"**Maximum Available Minutes**" is the total accumulated minutes of a provisioned AKS Cluster to the time Customer has initiated an action to stop or delete the cluster during an Applicable Period.
"**Downtime**" is the total accumulated minutes that are part of Maximum Available Minutes where a provisioned AKS Cluster has no connectivity to Kubernetes API Server.
"**Uptime Percentage**" is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for a given Microsoft Azure subscription. Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of AKS Clusters that don't have Availability Zones enabled:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%
< 95%	100%

Table of Contents / Definitions

Azure Lab Services

Additional Definitions:
"**Lab Virtual Machine(s)**" is defined as any virtual machine provisioned inside a lab in Azure Lab Services.
"**Lab Virtual Machine Connectivity**" is bi-directional network traffic between the Lab Virtual Machine and other IP addresses using TCP or UDP network protocols in which the Lab Virtual Machine is configured for allowed traffic. The IP addresses can be IP addresses within the same virtual network as the Lab Virtual Machine or public, routable IP addresses.
Uptime Calculation and Service Levels for Azure Lab Services
"**Minutes in the Applicable Period**" is the total number of minutes in a given Applicable Period.
"**Downtime**" is the total accumulated minutes that are part of Minutes in the Applicable Period that have no Lab Virtual Machine Connectivity.
"Uptime Percentage" is calculated by the percentage of Minutes in the Applicable Period for a given Microsoft Azure subscription, in which any Lab Virtual Machine had Downtime.
The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Minutes in the Applicable Period-Downtime}}{\text{Minutes in the Applicable Period}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of Lab Virtual Machines:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure Load Balancer

Additional Definitions:
"**Load Balanced Endpoint**" is an IP address and associated IP transport port definition.
"**Healthy Virtual Machine**" is a Virtual Machine which returns a Success Code for the health probe sent by the Azure Standard Load Balancer. The Virtual Machine must have Network Security Group rules permitting communication with the load balanced port.
"**Connectivity**" is bi-directional network traffic over supported IP transport protocols that can be sent and received from any IP address configured to allow traffic.
Uptime Calculation and Service Levels for Azure Load Balancer
"**Maximum Available Minutes**" is the total number of minutes that a given Azure Standard Load Balancer (serving two or more Healthy Virtual Machines) has been deployed by Customer in a Microsoft Azure subscription during an Applicable Period.
"**Downtime**" is the total number of minutes within Maximum Available Minutes during which the given Azure Standard Load Balancer is unavailable. A minute is considered unavailable if all Healthy Virtual Machines have no Connectivity through the Load Balanced Endpoint. Downtime does not include minutes resulting from SNAT port exhaustions.
"Uptime Percentage" for Azure Standard Load Balancer is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes multiplied by 100.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of Azure Load Balancer:

Uptime Percentage	Service Credit
< 99.99%	10%
< 99.9%	25%

Service Level Exceptions: No SLA is provided for Basic Load Balancer.

Table of Contents / Definitions

Azure Load Testing

Additional Definitions
"Deployment Minutes" is the total number of minutes that a given Azure Load Testing Service Resource has been deployed in Microsoft Azure during an Applicable Period.
"Maximum Available Minutes" is the sum of all Deployment Minutes across all Azure Load Testing Service Resources deployed by Customer in a given Microsoft Azure subscription during an Applicable Period.
"Downtime" is the sum of all Deployment Minutes, across all Azure Load Testing Service Resources deployed by Customer in a given Microsoft Azure subscription, during an Applicable Period during which the Load Test Resource is unavailable. A minute is considered unavailable for a given Service if all continuous HTTP requests to perform operations throughout the minute either result in an Error Code or do not return a response in 5 minutes.
"**Uptime Percentage**" The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of Azure Load Testing:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Log Analytics (Query Availability SLA)

Additional Definitions:
"Maximum Available Minutes" is the total number of minutes that a given Log Analytics Workspace has been deployed by Customer in a Microsoft Azure subscription during an Applicable Period.
"Downtime" is the total number of minutes within Maximum Available Minutes that data in a Log Analytics Workspace are unavailable. A minute is considered unavailable for a given Log Analytics Workspace during which no HTTP operations resulted in a Success Code.

" Query Availability Percentage" for a given Log Analytics Workspace calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes multiplied by 100.
Query Availability Percentage: The Query Availability Percentage is calculated using the following formula:

$$\frac{Maximum\,Available\,Minutes - Downtime}{Maximum\,Available\,Minutes} x100$$

Service Credit:

Query Availability Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Logic Apps

Additional Definitions:
"**Deployment Minutes**" is the total number of minutes that a given Logic App has been set to running in Microsoft Azure during an Applicable Period. Deployment Minutes is measured from when the Logic App was created or Customer initiated an action that would result in running the Logic App to the time Customer initiated an action that would result in stopping or deleting the Logic App.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Logic Apps deployed by Customer in a given Microsoft Azure subscription during an Applicable Period.
"**Downtime**" The total accumulated Deployment Minutes, across all Logic Apps deployed by Customer in a given Microsoft Azure subscription, during which the Logic App is unavailable. A minute is considered unavailable for a given Logic App when there is no connectivity between the Logic App and Microsoft's Internet gateway.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{Maximum\,Available\,Minutes - Downtime}{Maximum\,Available\,Minutes} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

As of 1 September 2024, Integration Services Environment is a retired Service feature, and its Support Window is expired. Logic Apps deployed into an Integration Services Environment no longer provides any Service Level guarantees, and by extension Service Credits, for performance or availability issues related to Logic Apps running on Integration Services Environment.

Table of Contents / Definitions

Azure Machine Learning

Uptime Calculation and Service Levels for Machine Learning Real Time Scoring
"**Total Transaction Attempts**" is the total number of API requests by Customer during an Applicable Period for a given Microsoft Azure subscription.
"**Failed Transactions**" is the set of all requests within Total Transaction Attempts that either return an Error Code or an HTTP 4xx status code or fail to return a Success Code within 600 seconds.
"**Uptime Percentage**" is calculated as Total Transaction Attempts less Failed Transactions divided by Total Transaction Attempts in an Applicable Period for a given Microsoft Azure subscription. Uptime Percentage is represented by the following formula:

$$\frac{\text{Total Transaction Attempts-Failed Transactions}}{\text{Total Transaction Attempts}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Machine Learning Realtime Scoring.

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Uptime Calculation and Service Levels for Machine Learning compute management
"**Total Transaction Attempts**" is the total number of API requests by Customer during an Applicable Period for a given Microsoft Azure subscription.
"**Failed Transactions**" is the set of all requests within Total Requests that either return an Error Code or an HTTP 408 status code or fail to return a Success Code within 30 seconds.
"**Uptime Percentage**" is calculated as Total Transaction Attempts less Failed Transactions divided by Total Transaction Attempts in an Applicable Period for a given Microsoft Azure subscription. Uptime Percentage is represented by the following formula:

$$\frac{\text{Total Transaction Attempts-Failed Transactions}}{\text{Total Transaction Attempts}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Machine Learning Management Plane Operations.

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure Managed Grafana

Additional Definitions:
"**Workspace**" is a deployment of one or more Grafana servers.
Monthly Uptime Calculation and Service Levels for Azure Managed Grafana
"**Deployment Minutes**" is the total number of minutes that a given Workspace has been running in Microsoft Azure during a billing month.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Workspaces deployed by Customer in a given Microsoft Azure subscription during a billing month.
"**Downtime**" is the sum of all Deployment Minutes, across all Workspaces deployed by Customer in a given Microsoft Azure subscription, during a billing month during which the Workspace is unavailable. A minute is considered unavailable for a given Workspace if all continuous HTTP requests to perform operations throughout the minute either result in an Error Code or do not return a response.
"**Monthly Uptime Percentage**" for the Azure Managed Grafana is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes. Monthly Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes - Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Azure Managed Grafana:

Monthly Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure Managed Instance for Apache Cassandra

Additional Definitions:
"**Availability Zone**" is a fault-isolated area within an Azure region, providing redundant power, cooling, and networking.
"**Cassandra Data Center**" refers to three or more Nodes deployed in a region using Azure Managed Instance for Apache Cassandra with a common configuration and Cassandra Data Center name.
"**Node**" refers to a single virtual machine deployed in a region using Azure Managed Instance for Apache Cassandra.
"**Compliant Networking Configuration**" means the full set of required configurations of the Microsoft Azure Virtual Network hosting the Cassandra Data Center, including Microsoft Azure Network Security Group inbound security rules and mandatory Microsoft Azure User Defined Routes of Microsoft Azure Virtual Network Subnet hosting the Cassandra Data Center, allowing uninterrupted flow of the management traffic and allowing data traffic to the dedicated gateway placed in the Microsoft Azure Virtual Network Subnet hosting the Cassandra Data Center.
Uptime Calculation and Service Levels for Azure Managed Instance for Cassandra Data Centers

"**Maximum Available Minutes**" is the total accumulated minutes during a billing month for all Cassandra Data Centers that have three or more Nodes deployed. Maximum Available Minutes is measured from when at least three Nodes in the same Cassandra Data Center have all been started resultant from action initiated by Customer to the time Customer has initiated an action that would result in stopping or deleting the Cassandra Data Center.
"**Downtime**" is the total accumulated minutes that are part of Maximum Available Minutes that continuously have no connectivity to a quorum of Nodes in the Cassandra Data Center in a region.
"**Uptime Percentage**" for Cassandra Data Centers is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for a given Microsoft Azure subscription. Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes - Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of a Cassandra Data Center with Compliant Network Configuration:

Uptime Percentage	Service Credit
< 99.95%	10%
< 99%	25%

For a Casandra Data Center deployed with Availability Zone support enabled in an Azure region that supports Availability Zones, the following Service Levels and Service Credits are applicable to Customer's use of a Cassandra Data Center with Compliant Network Configuration:

Uptime Percentage	Service Credit
< 99.99%	10%
< 99%	25%

For a Cassandra Cluster deployed with data centers in multiple regions and Availability Zone support enabled in an Azure region that supports Availability Zones, the following Service Levels and Service Credits are applicable to Customer's use of a Cassandra Data Center with Compliant Network Configuration:

Multiple Write Locations Availability Percentage	Service Credit
< 99.999%	10%
< 99%	25%

For a Casandra Data Center deployed with Single Region and Availability Zone is NOT enabled

Availability Percentage	Service Credit
< 99.995%	10%
< 99%	25%

For a Casandra Data Center deployed in a Single Region with Availability Zone enabled

Availability Percentage (SR-AZ)	Service Credit
< 99.995%	10%
< 99%	25%

Table of Contents / Definitions

Azure Maps

Additional Definitions:
"**Total Transaction Attempts**" is the total number of authenticated API requests made by Customer for a given Azure Map API during an Applicable Period in a given Microsoft Azure subscription. Total Transaction Attempts do not include API requests that return an Error Code that are continuously repeated within a five-minute window after the first Error Code is received.
"**Failed Transactions**" is the set of all requests within Total Transaction Attempts that result in an Error Code or otherwise do not return a Success Code within 60 seconds after receipt by the Service.
"Uptime Percentage" for a given Azure Map API is calculated as Total Transaction Attempts less Failed Transactions divided by Total Transaction Attempts multiplied by 100.
The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Total Transaction Attempts - Failed Transactions}}{\text{Total Transaction Attempts}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of Azure Maps API:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Media Services

Additional Definitions:
"**Allocated Egress Bandwidth**" is the amount of bandwidth configured by Customer in the Management Portal for a Media Service. Allocated Egress Bandwidth may be labeled "Streaming Units" or a similar name in the Management Portal.
"**Channel**" means an end point within a Media Service that is configured to receive media data.
"**Encoding**" means the processing of media files per subscription as configured in the Media Services Tasks.
"**Indexer Task**" means a Media Services Task that is configured to extract the speech content from an MP3 input file with a minimum five-minute duration.
"**Media Reserved Unit**" means reserved units purchased by the customer in an Azure Media Services account.
"**Media Service**" means an Azure Media Services account, created in the Management Portal, associated with Customer's Microsoft Azure subscription. Each Microsoft Azure subscription may have more than one associated Media Service.
"**Media Service Request**" means a request issued to Customer's Media Service.
"**Media Services Task**" means an individual operation of media processing work as configured by Customer. Media processing operations involve encoding and converting media files.
"**Streaming Unit**" means a unit of reserved egress capacity purchased by Customer for a Media Service.
"**Valid Key Requests**" are all requests made to the Content Protection Service for existing content keys in a Customer's Media Service.
"**Valid Media Services Requests**" are all qualifying Media Service Requests for existing media content in a customer's Azure Storage account associated with its Media Service when at least one Streaming Unit has been purchased and allocated to that Media Service. Valid Media Services Requests do not include Media Service Requests for which total throughput exceeds 80% of the Allocated Bandwidth.

Uptime Calculation and Service Levels for Encoding Service
"**Total Transaction Attempts**" is the total number of authenticated REST API requests with respect to a Media Service made by Customer during an Applicable Period for a subscription. Total Transaction Attempts does not include REST API requests that return an Error Code that are continuously repeated within a five-minute window after the first Error Code is received.
"**Failed Transactions**" is the set of all requests within Total Transaction Attempts that do not return a Success Code within 30 seconds from Microsoft's receipt of the request.
"Uptime Percentage" for the Azure Media Services Encoding Service is calculated as Total Transaction Attempts less Failed Transactions divided by Total Transaction Attempts in an Applicable Period for a given Microsoft Azure subscription.
The Uptime Percentage is calculated using the following formula:

$$\frac{Total\ Transaction\ Attempts - Failed\ Transactions}{Total\ Transaction\ Attempts}\ x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Azure Media Services Encoding Service:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Uptime Calculation and Service Levels for Media Indexer
Additional Definitions:
"**Total Transaction Attempts**" is the total number of Indexer Tasks attempted to be executed using an available Media Reserved Unit by Customer during an Applicable Period for a subscription.
"**Failed Transactions**" is the set of Indexer Tasks within Total Transaction Attempts that either, a) do not complete within a time period that is 3 times the duration of the input file, or b) do not start processing within 5 minutes of the time that a Media Reserved Unit becomes available for use by the Indexer Task.

"Uptime Percentage" for the Media Indexer is calculated as Total Transaction Attempts less Failed Transactions divided by Total Transaction Attempts in an Applicable Period for a given Microsoft Azure subscription.
The Uptime Percentage is calculated using the following formula:

$$\frac{Total\ Transaction\ Attempts - Failed\ Transactions}{Total\ Transaction\ Attempts} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Media Indexer:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Uptime Calculation and Service Levels for Streaming Service
Additional Definitions:
"**Deployment Minutes**" is the total number of minutes that a given Streaming Unit has been purchased and allocated to a Media Service during an Applicable Period.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Streaming Units purchased and allocated to a Media Service during an Applicable Period.
Downtime: The total accumulated Deployment Minutes when the Streaming Service is unavailable. A minute is considered unavailable for a given Streaming Unit if all continuous Valid Media Service Requests made to the Streaming Unit throughout the minute result in an Error Code.
"Uptime Percentage" for the Azure Media Services Streaming Service is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for a given Microsoft Azure subscription.
The Uptime Percentage is calculated using the following formula:

$$\frac{Maximum\ Available Minutes - Downtime}{Maximum\ Available Minutes} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Azure Media Services On-Demand Streaming Service:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Uptime Calculation and Service Levels for Video Indexer service
Additional Definitions:
"**Total Transaction Attempts**" is the total number of authenticated Video Indexer API requests made by Customer during an Applicable Period for a subscription. Total Transaction Attempts do not include Video Indexer API requests that return an Error Code that are continuously repeated within a five-minute window after the first Error Code is received, or Upload POST requests that send the file as byte array content.
"**Failed Transactions**" is the set of all requests within Total Transaction Attempts that return an Error Code, or do not send a response within 360 seconds from the completion of client sending the request.
"Uptime Percentage" for Video Indexer Service is calculated as Total Transaction Attempts less Failed Transactions divided by Total Transaction Attempts in an Applicable Period for a given Microsoft Azure subscription.
The Uptime Percentage is calculated using the following formula:

$$\frac{Total\ Transaction\ Attempts - Failed\ Transactions}{Total\ Transaction\ Attempts} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Azure Video Indexer Service:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Uptime Calculation and Service Levels for Live Channels
Additional Definitions:
"**Deployment Minutes**" is the total number of minutes that a given Channel has been purchased and allocated to a Media Service and is in a running state during an Applicable Period.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Channels purchased and allocated to a Media Service during an Applicable Period.

Downtime: The total accumulated Deployment Minutes when the Live Channels Service is unavailable. A minute is considered unavailable for a given Channel if the Channel has no External Connectivity during the minute.
"Uptime Percentage" for the Live Channels Service is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for a given Azure subscription.
The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Azure Media Services Live Channels Service:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Uptime Calculation and Service Levels for Content Protection Service
Additional Definitions
"**Total Transaction Attempts**" are all Valid Key Requests made by you during an Applicable Period for a given Azure subscription.
"**Failed Transactions**" are all Valid Key Requests included in Total Transaction Attempts that result in an Error Code or otherwise do not return a Success Code within 30 seconds after receipt by the Content Protection Service.
"Uptime Percentage" for Azure Media Services is calculated as Total Transaction Attempts less Failed Transactions divided by Total Transaction Attempts in an Applicable Period for a given Microsoft Azure subscription.
The Uptime Percentage is calculated using the following formula:

$$\frac{Total\ Transaction\ Attempts - Failed\ Transactions}{Total\ Transaction\ Attempts} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Azure Media Services Content Protection Service:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

MedTech service

A MedTech service is considered available within a one-minute window if it has successfully read from the configured data source or is available to read from a data source once properly configured.
Uptime Calculation
"**Uptime Percentage**" for the MedTech service is calculated as Total Minutes Available less Minutes Unavailable divided by Total Minutes the MedTech service was active. Uptime Percentage is represented by the following formula:

$$\frac{Total\ Minutes\ Available - Failed\ Minutes}{Total\ Minutes} x100$$

The following Service Levels and Service Credits are applicable to the MedTech service:
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Microsoft Fabric

Additional Definitions
Capacity: Capacity is a dedicated set of resources that is available at a given time to be used. Capacity defines the ability of a resource to perform an activity or to produce output. Different items consume different capacity at a

certain time. Fabric offers capacity through the Fabric SKU and Trials. For more information, see What is capacity (https://learn.microsoft.com/fabric/enterprise/licenses#capacity)?
Maximum Available Minutes: The sum of all minutes that a given, individual Capacity has been instantiated during an Applicable Period for a given tenant.
Downtime Minutes*: The total accumulated minutes in an Applicable Period for a given Capacity, after its creation, or before it is deprovisioned when the Capacity is unable to be utilized in all applicable Fabric features listed below:
Power BI – redirect to the Power BI section
View: View Power BI Dashboards, Reports, and Apps in the service.
Dataset Refresh: Schedule or manually trigger refresh operation and expect those operations to complete within expected timeframes considering all conditions that might impact refresh speeds (e.g., size of dataset).
Access Power BI Portal: Access and use the Power BI Portal within expected timeframes considering network conditions and limitations local to the customer environment or external to Microsoft.
Data Factory
Dataflow Gen2 Refresh: Schedule or manually trigger refresh operation and expect those operations to complete.
Open Pipeline: Open Pipelines in the service.
Data Engineering
Open Lakehouse: Open and view a Lakehouse in the service.
Open Notebook: Open and view a Notebook in the service.
Open Spark Job Definition: Open and view a Spark Job Definition in the service.
Open API for GraphQL: Open and view a GraphQL API in the service.
Data Science
Open ML model: Open and view ML models in the service.
Open Experiment: Open and view Experiments in the service.
Data Warehouse
Open Database: Open and view a Data Warehouse database in the service.
Real-Time Analytics
Open Eventstream: Open and view Eventstreams in the service.
Open KQL Database: Open and view a KQL database in the service.
OneLake
OneLake read transactions: Any read operations to OneLake DFS APIs.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

*Downtime Minutes does not include any period of time where the inability to use Fabric features within a given Capacity was due to Microsoft Fabric throttling policy.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Microsoft Genomics

Additional Definitions:
"**Maximum Available Minutes**" is the total accumulated minutes for all Microsoft Genomics accounts created by Customer and active during an Applicable Period for a given Microsoft Azure Subscription.
"**Downtime**" is the total number of minutes within Maximum Available Minutes during which Microsoft Genomics is unavailable. A minute is considered unavailable if all continuous attempts to send authenticated Genomics service REST API requests throughout the minute either return an Error Code or do not respond with an acknowledgement within the minute.
"**Uptime Percentage**" for Microsoft Genomics is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%

< 99%	25%

Table of Contents / Definitions

Microsoft Sentinel

Additional Definitions:
"**Maximum Available Minutes**" is the total number of minutes that a given Microsoft Sentinel has been deployed by Customer in a Microsoft Azure subscription during an Applicable Period.
"**Downtime**" is the total number of minutes within Maximum Available Minutes that data in Microsoft Sentinel are unavailable. A minute is considered unavailable for a given Azure Sentinel during which no HTTP operations resulted in a Success Code.
"**Query Availability Percentage**" for a given Microsoft Sentinel calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes multiplied by 100.
Query Availability Percentage: The Query Availability Percentage is calculated using the following formula:

Service Credit:

Query Availability Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Mobile Services

Additional Definitions:
"**Failed Transactions**" include any API calls included in Total Transaction Attempts that result in either an Error Code or do not return a Success Code.
"**Total Transaction Attempts**" are the total accumulated API calls made to the Azure Mobile Services during an Applicable Period for a given Microsoft Azure subscription for which the Azure Mobile Services are running.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{Total\ Transaction\ Attempts - Failed\ Transactions}{Total\ Transaction\ Attempts}\ x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Service Level Exceptions: The Service Levels and Service Credits are applicable to your use of the Standard and Premium Mobile Services tiers.

Table of Contents / Definitions

Azure Monitor

Uptime Calculation and Service Levels for the Azure Monitor Alerts
Additional Definitions:
"**Alert Rule**" is a collection of signal criteria used to generate alerts using monitoring event data already available to Alert Service for analysis.
"**Maximum Available Minutes**" is the total number of minutes which Alert Rule(s) are deployed by Customer in a given Microsoft Azure subscription during an Applicable Period.
"**Downtime**" is the total number of minutes within Maximum Available Minutes during which the Alert Rule is unavailable. A minute is considered unavailable for a given Alert Rule if all continuous attempts to analyze telemetry signals for resources defined within the Alert Rule throughout the minute either return an Error Code or do not result in a Success Code within five minutes from scheduled Alert Rule start time.
"**Uptime Percentage**" is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes multiplied by 100.
Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of Azure Monitor Alerts

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Uptime Calculation and Service Levels for the Azure Monitor Notification Delivery
Additional Definitions:
"**Action Group**" is a collection of actions which defines preferred notification delivery methods.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes which Action Groups are deployed by Customer in a given Microsoft Azure subscription during an Applicable Period.
Downtime: is the total number of minutes within Maximum Available Minutes during which the Action Group is unavailable. A minute is considered unavailable for a given Action Group if all continuous attempts to send alerts or perform registration management operations with respect to the Action Group throughout the minute either return an Error Code or do not result in a Success Code within five minutes.
Uptime Percentage: is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for a given Microsoft Azure subscription.

Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Levels and Service Credits:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Also refer to Log Analytics and Application Insights.
Monthly Uptime Calculation and Service Levels for the Azure Monitor managed service for Prometheus
Additional Definitions:
"Maximum Available Minutes" is the total number of minutes that a given Azure Monitor workspace has been deployed by Customer in a given Microsoft Azure subscription during a billing month.
Downtime: is the total number of minutes within Maximum Available Minutes that data in a given Azure Monitor workspace are unavailable. A minute is considered unavailable for a given Azure Monitor workspace if all continuous attempts to retrieve Prometheus metric data throughout the minute result in either an Error Code or do not return any HTTP code within 60 seconds.
Monthly Query Availability Percentage: for a given Azure Monitor workspace is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes multiplied by 100.
Monthly Query Availability Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Levels and Service Credits:

Monthly Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure NetApp Files

Additional Definitions
"**Volume**" is a logical storage resource in Azure NetApp Files that contains a file system and is used to store data.
"**Volume Connectivity**" is bi-directional network traffic between the Volume and other IP addresses using TCP or UDP network protocols in which the Volume is configured for allowed traffic.
"**Maximum Available Minutes**" is the total number of minutes which a Volume is deployed by Customer in a given Microsoft Azure subscription during an Applicable Period.

"**Downtime**" is the total accumulated minutes that are part of Maximum Available Minutes that have no Volume Connectivity in the Azure region.
"**Uptime Percentage**" The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit

Uptime Percentage	Service Credit
< 99.99%	10%
< 99%	25%

Table of Contents / Definitions

Network Watcher

Additional Definitions:
"**Network Diagnostic Tools**" is a collection of network diagnostic and topology tools.
"**Maximum Diagnostic Checks**" is the total number of diagnostic actions performed by the Network Diagnostic Tool as configured by Customer in an Applicable Period for a given Microsoft Azure subscription.
"**Failed Diagnostic Checks**" is the total number of diagnostic actions within Maximum Diagnostic Checks that returns an Error Code or does not return a response within the Maximum Processing Time documented in the table below.

Diagnostic Tool	Maximum Processing Time
IPFlow Verify NextHop Packet Capture Security Group View Topology Connection Monitor Connection Monitor (classic)	2 minutes
VPN Troubleshoot	10 minutes

"**Uptime Percentage**" is calculated by using the following formula:

$$\frac{\text{Maximum Diagnostic Checks - Failed Diagnostic Checks}}{\text{Maximum Diagnostic Checks}} x100$$

Service Levels:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Notification Hubs

Additional Definitions:
"**Deployment Minutes**" is the total number of minutes that a given Notification Hub has been deployed in Microsoft Azure during an Applicable Period.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Notification Hubs deployed by you in a given Microsoft Azure subscription under the Basic or Standard Notification Hubs tiers during an Applicable Period.
Downtime: The total accumulated Deployment Minutes, across all Notification Hubs deployed by you in a given Microsoft Azure subscription under the Basic or Standard Notification Hubs tiers, during which the Notification Hub is unavailable. A minute is considered unavailable for a given Notification Hub if all continuous attempts to send notifications or perform registration management operations with respect to the Notification Hub throughout the minute either return an Error Code or do not result in a Success Code within five minutes.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Service Level Exceptions: The Service Levels and Service Credits are applicable to your use of the Basic and Standard Notification Hubs tiers.

Table of Contents / Definitions

On Demand Capacity Reservations for Azure Virtual Machines

Additional Definitions
"**Availability Zone**" is a fault-isolated area within an Azure region, providing redundant power, cooling, and networking.
"**Capacity**" is a property of an On Demand Capacity Reservation that specifies the quantity of Virtual Machine instances reserved
"**On Demand Capacity Reservation**" is an object created in an Azure subscription to express a quantity of reserved capacity for a specific Virtual Machine instance type in a specific location.
"**Virtual Machine**" refers to persistent instance types that can be deployed individually, or as part of a Virtual Machine Scale Set, in a multi-tenant environment in Azure.
"**Virtual Machines Allocated**" is a property of an On Demand Capacity Reservation; this refers to the list of Virtual Machines allocated to the On Demand Capacity Reservation.
"**Supported Deployment**" is a Virtual Machine deployment that matches the location, including Availability Zone if defined, uses the exact Virtual Machine size of an existing On Demand Capacity Reservation, and complies with the usage documentation of the feature.
"**Reserved Unit**" is exactly one instance of an On Demand Capacity Reservation. For example, if an On Demand Capacity Reservation specifies a Capacity of 10 Virtual Machines, then there are 10 Reserved Units.
"**Unused Capacity Reservation**" is an On Demand Capacity Reservation with the count of Virtual Machines Allocated less than the Capacity.
"**Not Available for Deployment**" is defined as any Supported Deployment configured to consume an existing Unused Capacity Reservation that meets two conditions:

i. Receives an error that specifies lack of Virtual Machine capacity. A Virtual Machine deployment failure for other error types or due to lack of capacity for disks or any other Azure resource does not meet this requirement; and
ii. On Demand Capacity Reservation continues to meet the definition of Unused Capacity Reservation (e.g., other Virtual Machine(s) did not already consume the Unused Capacity Reservation).

Uptime Calculation and Service Levels for On Demand Capacity Reservation
"**Minutes not Available**" is defined as minutes an Unused Capacity Reservation is Not Available for Deployment. From the time a Not Available for Deployment condition occurs, Minutes not Available will accumulate until (a) a subsequent Supported Deployment succeeds, (b) another Not Available for Deployment condition results from another Supported Deployment attempt, or (c) 15 minutes has elapsed. If 15 minutes elapses with no Supported Deployment attempted, then Minutes not Available will resume accumulation upon a subsequent Not Available for Deployment condition.

The Minutes not Available will accumulate for each Reserved Unit that cannot be used. If one Reserved Unit becomes used while another remains unused, then Minutes not Available will continue to accumulate only for the unused Reserved Unit.

See example calculation in the feature documentation.

"**Downtime**" is the total accumulated Minutes not Available that are part of minutes in a given month calculated per Reserved Unit.
"**Uptime Percentage**" for each Reserved Unit is calculated by the percentage of Minutes in the Applicable Period in which a Reserved Unit had Downtime.

$$\frac{\text{Minutes in an Applicable Period-Downtime}}{\text{Minutes in an Applicable Period}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of each Reserved Unit in an On Demand Capacity Reservation. The Service Credits are issued based on the cost of each Reserved Unit, not the overall cost of the On Demand Capacity Reservation.

Uptime Percentage	Service Credit
< 99.9%	10%

Uptime Percentage	Service Credit
< 99%	25%
< 95%	100%

Table of Contents / Definitions

Azure AI Foundry Models

Uptime Calculation and Service Levels for Azure AI Foundry Models
Additional Definitions:
"Azure AI Foundry Resource" refers to an Azure resource for Azure AI Foundry created in an Azure region in a Microsoft Azure subscription.
"Request" is an API call to a model endpoint in an Azure AI Foundry Resource.
"Average Error Rate" for an Applicable Period is the sum of Error Rates for each minute in the Applicable Period divided by the total number of minutes in the Applicable Period.
"Error Rate" is the total number of Requests that return an Error Code divided by the total number of Requests during a minute. If the total number of Requests in a given minute interval is zero, the error rate for that interval is 0%.
"Uptime Percentage" is represented by the following formula

$$100\% - Average\,Error\,Rate$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Service Level Exceptions: The Service Levels and Service Credits are applicable to your use of the Standard tier of the Azure AI Foundry Models and only applicable to Azure Direct Models. The Developer tier of the Azure AI Foundry Models Service is not covered by this SLA.

Latency Calculation and Service Levels for Azure AI Foundry Models with Provisioned Throughput
Additional Definitions
"**Azure AI Foundry Models Provisioned Throughput Deployment**" is an Azure AI Foundry Model deployed using provisioned throughput for an applicable model within an Azure AI Foundry Resource.
"**Maximum Request Size**" is the maximum number of input and output tokens that can be included in a given request for a specified Azure AI Foundry model, as defined by the product documentation.
"**N**" is the number of Successful Requests made to an Azure AI Foundry Models Provisioned Throughput Deployment with a maximum request size that is less than or equal to the limit provided for the specified Azure OpenAI model in a given minute.
"**Tokens Per Second**" is a latency measurement that enumerates the generation speed for a given Azure AI Foundry Model response. The total tokens generated is divided by the time to generate the tokens. Generated tokens are counted as all tokens generated after the first token.
"**SLA Latency Target Value**" is the latency target in Tokens Per Second for a given model as defined by the product documentation.
"**S**" are the response times in N that meet the following criteria and are determined as follows:

- Tokens Per Second for calls generating more tokens than the SLA Latency Target Value OR taking more than 1 second to generate all response
- SLA Latency Target Value for calls generating less than or equal to the SLA Latency Target Value AND taking less than 1 second to generate all response tokens
- 0 for all calls generating 1 or fewer tokens

"**Average Tokens Per Second**" is the average of the Tokens Per Second for all S in a given minute.
"**Excessive Latency Minutes**": is the total number of one-minute intervals during which the Average Tokens Per Second is less than the SLA Latency Target Value. If N is zero, the Excessive Latency Minutes for that interval is 0.
"**Average Excessive Latency Rate**" for an Applicable Period is the sum of Excessive Latency Minutes divided by the total number of minutes in the Applicable Period.
"**Average Attainment Percentage**" for a given Azure AI Foundry Models Provisioned Throughput Deployment is calculated by subtracting from 100% the Average Excessive Latency Rate for a given Azure AI Foundry Resource in an Applicable Period.

Time Between Token Latency Attainment Percentage is represented by the following formula:

$$100\% - Average\,Excessive\,Latency\,Rate$$

Service Credit:

Average Attainment Percentage	Service Credit
< 99%	10%

Service Level Exceptions: The Service Levels and Service Credits are applicable to your use of the Standard tier of the Azure AI Foundry Models and only applicable to Azure Direct Models where the SLA Latency Target Value is defined in the product documentation. The Developer tier of the Azure AI Foundry Models Service is not covered by this SLA.

Table of Contents / Definitions

Object Replication - Priority Replication

This SLA applies to any Replication Policy where the Source Storage Account and Destination Storage Account are located within the same continent and are eligible for Priority Replication.

Additional Definitions:

"**Destination Container**" is the container which operations are asynchronously replicated to within an Object Replication Policy.

"**Destination Storage Account**" is a storage account where operations are asynchronously replicated to, based upon a Replication Policy a user configures with Object Replication on their Source Storage Account.

"**Exceeded Lag Operation**" is any operation (data, metadata, properties or deletes) that is pending replication longer than 15 minutes or takes longer than 15 minutes to be replicated from the Source Storage Account to the Destination Storage Account during the billing month.

"**Object Replication**" is a storage account feature that asynchronously copies block blobs from a Source Storage Account to a Destination Storage Account. The block blobs within a container are replicated according to Replication Rules a user configures. The contents of the blob, any versions associated with the blob, and the blob's metadata, properties and deletes are all copied from the Source Container to the Destination Container.

"**Replication Lag**" indicates the amount of time that has elapsed between when an operation is performed on the Source Storage Account and when the operation is replicated to the Destination Storage Account.

"**Replication Policy**" specifies the Source Storage Account and the Destination Storage Account. A Replication Policy includes one or more rules that specify a Source Container and a Destination Container and indicate which block blobs in the Source Container will be replicated.

"**Replication Rule**" specifies how Azure Storage will replicate operations from a Source Container to a Destination Container.

"**Source Container**" is the container an operation is originally uploaded to within a Replication Policy.

"**Source Storage Account**" is a storage account where Replication Policies and Replication Rules are created. Operations are performed on this storage account, then asynchronously replicated to the Destination Storage Account.

"**Total Operations**" is the total count of operations that are replicated or pending replication by Object Replication for the applicable period during the billing month.

"Monthly **Conformance Percentage**": The Monthly Conformance Percentage is calculated using the following formula:

$$\frac{Total\,Operations - Exceeded\,Lag\,Operations}{Total\,Operations}\,x100$$

Service Credit:

Conformance Percentage	Service Credit
Less than 99.0% and greater than or equal to 95.0%	10%
Less than 95.0% and greater than or equal to 90.0%	25%
<90.00%	100%

Service Level Exceptions: This SLA does not apply to:

1. Storage accounts that do not have Object Replication Priority Replication enabled,
2. Objects whose size exceed 5 gigabytes (GB),
3. Objects that are modified more frequently than 10 times per second,
4. Object Replication Policies where the Source Storage Account and Destination Storage Account are not within the same continent,
5. Storage accounts that are (a) larger than 5 petabytes (PB) or (b) have more than 10 billion blobs and
6. During time periods where (a) your storage account or Replication Policy data transfer rate exceeds 1 gigabit per second (Gbps) and the resulting back log of writes are being replicated, (b) your storage account or Replication Policy exceeds 1,000 PUT or DELETE operations per second and the resulting back log of writes are being replicated, and (c) existing blob replication is pending following a recent Replication Policy creation or update. Existing blob replication is estimated to progress at 100 TB per day on average but may experience reduced velocity when blobs with many versions are present.

Table of Contents / Definitions

Azure Operator Service Manager

Additional Definitions:
"Applicable Service Fees" means the total fees actually paid by you for a Service that are applied to the Applicable Period in which a Service Credit is owed.
"Service Level" means the performance metric(s) set forth in this SLA that Microsoft agrees to meet in the delivery of the Services.
"Maximum Available Minutes" is the total accumulated minutes during an Applicable Period in which at least one Azure Operator Service Manager Site Network Service resource has been deployed in a Microsoft Azure Subscription.
"Downtime" A minute is considered downtime if all continuous attempts to create, update or view resources of the service throughout the minute either return an Error Code or do not result in a Success Code within two minutes.
"Uptime Percentage" for Azure Operator Service Manager is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period. Uptime Percentage is calculated per region and is represented by the following formula:

Monthly Uptime %= (Maximum Available Minutes−Downtime)/Maximum Available Minutes x 100

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Service Level Exceptions:
Performance or availability issues that result from

- Your failure to adhere to any required configurations, use supported platforms, follow any policies for acceptable use, or your use of the Service in a manner inconsistent with the features and functionality of the Service (for example, attempts to perform operations that are not supported) or inconsistent with our published guidance.
- Your attempts to perform operations that exceed prescribed quotas or that resulted from our throttling of suspected abusive behavior.

Table of Contents / Definitions

Azure Orbital Ground Station

Additional Definitions
"**Azure Orbital Ground Station Service**" is a fully managed ground station service delivering low latency connectivity from customer satellites in orbit to the Microsoft Azure cloud. For purposes of this SLA, the service comprises connectivity only to Microsoft owned and operated ground stations and not ground stations owned or operated by third-party ground station providers that customers configure to work with the Azure Orbital Ground Station Service.
"**Scheduled Contact**" is the time when a connection between a Microsoft owned and operated Azure Orbital Ground Station and customer's satellite has been requested by customer (via either Azure Portal or API), and confirmed as scheduled by Microsoft (i.e., the status of the requested contact shows "Scheduled" in the Azure Portal or API interface).
"**Downtime**" is the time during a Scheduled Contact when customer is unable to pass data end-to-end either to, or from the customer’s satellites to customer’s virtual network endpoint on Azure, in either case due to a failure specifically attributable to the Azure Orbital Ground Station Service , or to the cancellation of a previously confirmed Scheduled Contact (as indicated by the ContactsStatus field changing from “scheduled” to either “failed” or “providerCancelled”).

Downtime does not include any time during a Scheduled Contact when customer is unable to pass data end-to-end due to (i) any circumstances falling under the below Service Level Exceptions; or (ii) any circumstances that prevent the Azure Orbital Ground Station Service from confirming a requested contact as a Scheduled Contact. For purposes of the Azure Orbital Ground Station Service, Downtime does not include Scheduled Downtime attributable to maintenance or repairs to the Azure Orbital Ground Station Service sites.

Service Level Exceptions:

- Performance or availability issues that result from
  - Restrictions or limitations placed on customer’s spacecraft license to operate by a regulatory body or government agency.

- Use of services, hardware, or software not provided by Microsoft, including, but not limited to, issues resulting from inadequate bandwidth or related to third-party software or services.
- Failure to adhere to any required configurations, use supported platforms, follow any policies for acceptable use, or use of Azure Orbital Ground Station Service in a manner inconsistent with its features and functionality (for example, attempts to perform operations that are not expressly supported) or inconsistent with our published documentation or guidance.

Successful Contact Percentage Calculation and Service Levels

"**Successful Contact Percentage**" is calculated as total Scheduled Contact minutes less total Downtime minutes divided by total Scheduled Contact minutes. It is represented by the following formula:

$$\frac{\text{Total Scheduled Contact Minutes-Downtime Minutes}}{\text{Total Scheduled Contact Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Azure Orbital Ground Station Service for the Applicable Service Fees:

Successful Contact Percentage	Service Credit
< 99.9%	10%
< 98%	25%
< 95%	100%

Table of Contents / Definitions

Microsoft Playwright Testing

Additional Definitions

"Deployment Minutes" is the total number of minutes that a given Microsoft Playwright Testing Service resource has been deployed in Microsoft Azure during an Applicable Period.

"Maximum Available Minutes" is the sum of all Deployment Minutes across all Microsoft Playwright Testing Service Resources deployed by customers in a given Microsoft Azure subscription during an Applicable Period.

"Downtime" is the sum of all Deployment Minutes, across all Microsoft Playwright Testing Service Resources deployed by Customer in a given Microsoft Azure subscription, during an Applicable Period during which the Microsoft Playwright Testing Resource is unavailable. A minute is considered unavailable for a given Service if all continuous HTTP requests to perform operations throughout the minute either result in an Error Code or do not return a response in 5 minutes.

"**Uptime Percentage**" The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of Microsoft Playwright Testing:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure Private Link

Additional Definitions

"**Azure Private Link Service**" is the reference to your own service that is enabled for Azure Private Link and deployed within your own virtual network.

"**Azure Private Endpoint**" is a network interface connecting your Azure Private Link enabled service to a private IP address in your virtual network.

Uptime Calculation

"**Maximum Available Minutes**" is the total accumulated minutes in an Applicable Period during which the Azure Private Link Service or Azure Private Endpoint have been deployed in a Microsoft Azure subscription.

"**Downtime**" is the total accumulated Maximum Available Minutes in an Applicable Period for a given Azure Private Link Service or Azure Private Endpoint during which the Azure Private Link Service or Azure Private Endpoint is unavailable. A given minute is considered unavailable if all attempts to connect through Azure Private Endpoint throughout the minute are unsuccessful.

"**Uptime Percentage**" The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Uptime Percentage	Service Credit
< 99.99%	10%
< 99%	25%

Table of Contents / Definitions

Microsoft Purview

Additional Definitions:
"**Total Requests**" is the set of all authenticated API requests, other than Excluded Requests, to perform Microsoft Purview operations during an Applicable Period for a given Microsoft Azure subscription.
"**Excluded Requests**" is the set of requests that result in an HTTP 4xx status code.
"**Failed Requests**" is the set of all requests within Total Requests that return an Error Code.
"**Uptime Percentage**" for the API calls made to the Microsoft Purview service is calculated as Total Requests less Failed Requests divided by Total Requests in an Applicable Period for a given Microsoft Azure subscription.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Total Requests-Failed Requests}}{\text{Total Requests}} x100$$

The following Service Credits are applicable to Customer's use of API calls within the Microsoft Purview* service:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

*The above Service Credits are only available to portions of Microsoft Purview that are offered on a subscription basis (formerly known as Azure Purview).

Table of Contents / Definitions

Azure Red Hat OpenShift

Additional Definitions
"**Maximum Available Minutes**" is the total accumulated minutes during an Applicable Period during which a given Azure Red Hat OpenShift cluster has been deployed in a Microsoft Azure subscription.
"**Downtime**" is the total accumulated Maximum Available Minutes during an Applicable Period in which a given Azure Red Hat OpenShift cluster API endpoint is unavailable. A given minute is considered unavailable if all attempts to connect to the cluster API endpoint throughout the minute are unsuccessful.
"**Uptime Percentage**" The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Uptime Percentage	Service Credit
< 99.95%	10%
< 99%	25%

Table of Contents / Definitions

Remote Rendering

Additional Definitions
"**Conversion**" refers to a process that transforms 3D models into the format required during a Rendering Session.
"**Rendering Session**" refers to an interaction with the Remote Rendering Service.
Uptime Calculation and Service Levels for Conversion REST API Transactions
"**Total Transaction Attempts**" is the total number of authenticated REST API requests for the Conversion functionality in the Azure Remote Rendering Service made by Customer during an Applicable Period for a subscription.

Total Transaction Attempts does not include REST API requests that return an Error Code that are continuously repeated within a five-minute window after the first Error Code is received.
"**Failed Transactions**" is the set of all requests within Total Transaction Attempts that returns an Error Code within 30 seconds from Microsoft's receipt of the request.
"**Uptime Percentage**" for the Azure Remote Rendering Service is calculated as Total Transaction Attempts less Failed Transactions divided by Total Transaction Attempts in an Applicable Period for a given Microsoft Azure subscription. Uptime Percentage is represented by the following formula:

$$\frac{Total\ Transaction\ Attempts - Failed\ Transactions}{Total\ Transaction\ Attempts} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Conversion functionality of the Azure Remote Rendering Service:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Uptime Calculation and Service Levels for Rendering Sessions
"**Deployment Minutes**" is the total number of minutes in a Rendering Session, measured from when a Rendering Session has been allocated due to an action initiated by Customer to the time the Customer has initiated an action that would result in stopping the session during an Applicable Period.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Rendering Session during an Applicable Period.
"**Downtime**" is the total accumulated Deployment Minutes when the Remote Rendering Service is unavailable. A minute is considered unavailable for a given Rendering Session if the Rendering Session has no External Connectivity during the minute.
"**Uptime Percentage**" for the Rendering Session is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for a given Azure subscription. Uptime Percentage is represented by the following formula:

$$\frac{Maximum\ Available Minutes - Downtime}{Maximum\ Available Minutes} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Rendering Sessions in Azure Remote Rendering Service:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure Route Server

Uptime Calculation
"**Maximum Available Minutes**" is the total accumulated minutes during an Applicable Period during which a given Azure Route Server has been deployed in a Microsoft Azure subscription.
"**Downtime**" is the total accumulated Maximum Available Minutes during which an Azure Route Server is unavailable. A minute is considered unavailable if all attempts to connect to the Azure Route Server within the minute are unsuccessful.
"**Uptime Percentage**" for a given Azure Route Server is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for a given Microsoft Azure subscription. Uptime Percentage is represented by the following formula:

$$\frac{Maximum\ Available Minutes - Downtime}{Maximum\ Available Minutes} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of each Azure Route Server:

Uptime Percentage	Service Credit
< 99.95%	10%
< 99%	25%

Table of Contents / Definitions

SAP HANA on Azure Large Instances

Additional Definitions:
"**Announced Single Instance Maintenance**" means periods of Downtime related to network, hardware, or Service maintenance or upgrades impacting Single Instances. We will publish notice or notify you at least five (5) days prior to the commencement of such Downtime.
"**High Availability Pair**" refers to two or more identical SAP HANA on Azure large instances deployed in the same region and configured by the customer for system replication at the application layer. Customer must declare the members of a High Availability Pair to Microsoft during the architecture design process.
"**SAP HANA on Azure Connectivity**" is bi-directional network traffic between the SAP HANA on Azure large instance and other IP addresses using TCP or UDP network protocols in which the instance is configured for allowed traffic. The IP addresses must be IP addresses on the Virtual Network of the associated Azure subscription.
"**Single Instance**" is defined as any single Microsoft SAP HANA on Azure Large Instance machine that is not deployed in an High Availability Pair.

Uptime Calculation and Service Levels for SAP HANA on Azure High Availability Pair

"**Maximum Available Minutes**" is the total accumulated minutes during an Applicable Period for all SAP HANA on Azure instances deployed in the same High Availability Pair. Maximum Available Minutes is measured from when two or more instances in the same High Availability Pair have both been started resultant from an action initiated by Customer to the time Customer has initiated an action that would result in stopping the instances.
"**Downtime**" is the total accumulated minutes that are part of Maximum Available Minutes that have no SAP HANA on Azure Connectivity.
Uptime Percentage: The Uptime Percentage for SAP HANA on Azure High Availability Pair is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit for SAP HANA on Azure High Availability Pair:

Uptime Percentage	Service Credit
< 99.99%	10%
< 99.9%	25%

Uptime Calculation and Service Levels for SAP HANA on Azure Single Instance

"**Maximum Available Minutes**" is the total accumulated minutes for all SAP HANA on Azure Single Instances deployed by Customer during an Applicable Period for a given Microsoft Azure subscription.
"**Downtime**" is the total accumulated minutes that are part of Maximum Available Minutes that have no SAP HANA on Azure Connectivity. Downtime excludes Announced Single Instance Maintenance.
Uptime Percentage: The Uptime Percentage for SAP HANA on Azure Single Instance is calculated using the following formula

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of SAP HANA on Azure Single Instances:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%
<95%	100%

Table of Contents / Definitions

Scheduler

Additional Definitions:
"**Maximum Available Minutes**" is the total number of minutes in an Applicable Period.
"**Planned Execution Time**" is a time at which a Scheduled Job is scheduled to begin executing.

"**Scheduled Job**" means an action specified by you to execute within Microsoft Azure according to a specified schedule.
Downtime: The total accumulated minutes in an Applicable Period during which one or more of your Scheduled Jobs is in a state of delayed execution. A given Scheduled Job is in a state of delayed execution if it has not begun executing after a Planned Execution Time, provided that such delayed execution time shall not be considered Downtime if the Scheduled Job begins executing within thirty (30) minutes after a Planned Execution Time.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Service-Bus

Additional Definitions:
"**Message**" refers to any user-defined content sent or received through Service Bus Relays, Queues, or Topics, using any protocol supported by Service Bus.
"Partitioned namespaces" enable messaging entities to be divided across multiple message brokers to increase overall throughput.
"**Availability Zone**" is a fault-isolated area within an Azure region, providing redundant power, cooling, and networking.
Uptime Calculation and Service Levels for Queues and Topics in all tiers deployed without partitioned namespaces
Additional Definitions:
"**Deployment Minutes**" is the total number of minutes that a given Queue or Topic has been deployed in Microsoft Azure during an Applicable Period.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Queues and Topics deployed by you in a given Microsoft Azure subscription during an Applicable Period.
Downtime: The total accumulated Deployment Minutes, across all Queues and Topics deployed by you in a given Microsoft Azure subscription, during which the Queue or Topic is unavailable. A minute is considered unavailable for a given Queue or Topic if all continuous attempts to send or receive Messages or perform other operations on the Queue or Topic throughout the minute either return an Error Code or do not result in a Success Code within five minutes.
Uptime Percentage: for Queues and Topics is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for a given Microsoft Azure subscription.
Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of Queues and Topics in all tiers deployed without partitioned namespaces:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Uptime Calculation and Service Levels for Queues and Topics in the Premium tier deployed with partitioned namespaces in regions with Availability Zone support
Additional Definitions:
"**Deployment Minutes**" is the total number of minutes that a given Queue or Topic has been deployed in Microsoft Azure during an Applicable Period.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Queues and Topics deployed by you in a given Microsoft Azure subscription during an Applicable Period.
Downtime: The total accumulated Deployment Minutes, across all Queues and Topics deployed by you in a given Microsoft Azure subscription, during which the Queue or Topic is unavailable. A minute is considered unavailable for a given Queue or Topic if all continuous attempts to send or receive Messages or perform other operations on the Queue or Topic throughout the minute either return an Error Code or do not result in a Success Code within five minutes.

Uptime Percentage: for Queues and Topics is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for a given Microsoft Azure subscription.
Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of Queues and Topics in the Premium tier deployed with partitioned namespaces in regions with Availability Zone support:

Uptime Percentage	Service Credit
< 99.99%	10%
< 99%	25%

Uptime Calculation and Service Levels for Relays
"**Deployment Minutes**" is the total number of minutes that a given Relay has been deployed in Microsoft Azure during an Applicable Period.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Relays deployed by Customer in a given Microsoft Azure subscription during an Applicable Period.
Downtime: Is the total accumulated Deployment Minutes, across all Relays deployed by Customer in a given Microsoft Azure subscription, during which the Relay is unavailable. A minute is considered unavailable for a given Relay if all continuous attempts to establish a connection to the Relay throughout the minute either return an Error Code or do not result in a Success Code within five minutes.
Uptime Percentage: The Uptime Percentage for Relays is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for a given Microsoft Azure subscription.
Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of Relays:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure SignalR Service

Additional Definitions:
"**Downtime**" is the total accumulated Maximum Available Minutes during an Applicable Period for the SignalR Service during which the SignalR Service is unavailable. A given minute is considered unavailable if all attempts to send SignalR Transactions throughout the minute either return an Error Code or do not result in a Success Code within one minute.
"**Maximum Available Minutes**" is the total number of minutes that the SignalR Service has been deployed by the Customer in a given Microsoft Azure subscription during an Applicable Period.
"**SignalR Service Endpoint**" is the host name from which the SignalR Service is accessed by servers or clients to perform SignalR Transactions.
"**SignalR Transactions**" is the set of transaction requests sent from client to server or from server to client through a SignalR Service Endpoint.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the SignalR Service Standard tiers.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Azure Site Recovery

Additional Definitions:
"**Failover**" is the process of transferring control, either simulated or actual, of a Protected Instance from a primary site to a secondary site.
"**On-Premises-to-Azure Failover**" is the Failover of a Protected Instance from a non-Azure primary site to an Azure secondary site.
"**Azure-to-Azure Failover**" is the Failover of a Protected Instance from an Azure primary site to an Azure secondary site.
"**Protected Instance**" refers to a virtual or physical machine configured for replication by the Site Recovery Service from a primary site to a secondary site. Protected Instances are enumerated in the Protected Items tab in the Recovery Services section of the Management Portal.

Uptime Calculation and Service Levels for Azure-to-Azure or On-Premises-to-Azure Failover
"**Failover Minutes**" is the total number of minutes in an Applicable Period during which a Failover of a Protected Instance configured for On-Premises-to-On-Premises replication has been attempted but not completed.
"**Maximum Available Minutes**" is the total number of minutes that a given Protected Instance has been configured for Azure-to-Azure or On-Premises-to-Azure replication by the Azure Site Recovery Service during an Applicable Period.
"**Protected Instance**" refers to a virtual or physical machine configured for replication by the Azure Site Recovery Service from a primary site to a secondary site. Protected Instances are enumerated in the Protected Items tab in the Recovery Services section of the Management Portal.
Downtime: Is the total accumulated Failover Minutes in which the Failover of a Protected Instance is unsuccessful due to unavailability of the Azure Site Recovery Service, provided that retries are continually attempted no less frequently than once every thirty minutes.
Uptime Percentage: for Azure-to-Azure or On-Premises-to-Azure Failover of a specific Protected Instance in a given Applicable Period is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes. The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of each Protected Instance within the Site Recovery Service for Azure-to-Azure or On-Premises-to-Azure Failover:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Recovery Time Objective and Service Levels for On-Premises-to-Azure Failover
"**Recovery Time Objective (RTO)**" means the period of time beginning when Customer initiates a Failover of a Protected Instance experiencing either a planned or unplanned outage for On-Premises-to-Azure replication to the time when the Protected Instance is running as a virtual machine in Microsoft Azure, excluding any time associated with manual action or the execution of Customer scripts.
"**Recovery Time Objective**": For a specific Protected Instance configured for On-Premises-to-Azure replication in a given Applicable Period is one hour.
The following Service Levels and Service Credits are applicable to Customer's use of each Protected Instance within the Site Recovery Service for On-Premises-to-Azure Failover.:

Recovery Time Objective	Service Credit
> 1 hour	100%

Recovery Time Objective and Service Levels for Azure-to-Azure Failover
"**Recovery Time Objective (RTO)**" means the period of time beginning when Customer initiates a Failover of a Protected Instance for Azure-to-Azure replication to the time when the Protected Instance is running as a virtual machine in secondary Azure region, excluding any time associated with manual action or the execution of Customer scripts.
"**Recovery Time Objective**" for a specific Protected Instance configured for Azure-to-Azure replication in a given Applicable Period is one hour.
The following Service Levels and Service Credits are applicable to Customer's use of each Protected Instance within the Site Recovery Service for Azure-to-Azure Failover:

Recovery Time Objective	Service Credit
>1 hour	100%

Note: Service credits for failover failure will not be applicable if compute capacity in the secondary region is not available.

Table of Contents / Definitions

Spatial Anchors

Additional Definitions
"**Total Transaction Attempts**" is the total number of authenticated API requests with respect to Azure Spatial Anchors made by Customer during an Applicable Period for a given Azure Spatial Anchors API. Total Transaction Attempts do not include API requests that return an Error Code that are continuously repeated within a five-minute window after the first Error Code is received.
"**Failed Transactions**" is the set of all requests to the Azure Spatial Anchors API within Total Transaction Attempts that return an Error Code.
Uptime Calculation
"**Uptime Percentage**" for Azure Spatial Anchors is calculated as Total Transaction Attempts less Failed Transactions divided by Total Transaction Attempts in an Applicable Period for a given Microsoft Azure subscription. Uptime Percentage is represented by the following formula:

$$\frac{Total\ Transaction\ Attempts - Failed\ Transactions}{Total\ Transaction\ Attempts} x100$$

The following Service Levels and Service Credits are applicable to Azure Spatial Anchors APIs:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure Spring Apps

Additional Definitions
"**App**" is a Spring Boot App deployed by Customer within Azure Spring Apps. Excluding Apps in the Basic Tier.
"**Spring Apps Service Runtime**" is a collection of Spring Apps components (e.g. Spring Apps Config Server, Spring Apps Registry) hosted by Microsoft.
Uptime Calculation and Service Levels for Azure Spring Apps
"**Deployment Minutes**" is the total number of minutes that a given App has been set to running in Microsoft Azure during an Applicable Period. Deployment Minutes is measured from when the App was created or Customer initiated an action that would result in running the App to the time Customer initiated an action that would result in stopping or deleting the App.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Apps deployed by Customer in a given Microsoft Azure subscription during an Applicable Period.
"**Downtime**" is the sum of all Deployment Minutes, across all Apps deployed by Customer in a given Microsoft Azure subscription, during an Applicable Period during which the App is unavailable. A minute is considered unavailable for a given App if all continuous attempts to connect between the App and Microsoft's Internet gateway or the Azure Spring Apps Service Runtime throughout the minute result in either an Error Code or do not return a Success Code within five minutes.
"**Uptime Percentage**" The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Standard Tier:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

The following Service Levels and Service Credits are applicable to Enterprise Tier:

Uptime Percentage	Service Credit
< 99.95%	10%
< 99%	25%

Azure SQL Database

Additional Definitions:
"**Availability Zone**" is a fault-isolated area within an Azure region, providing redundant power, cooling, and networking.
"**Database**" means any Microsoft Azure SQL Database created in any of the Service tiers and deployed either as a single database or in an Elastic Pool.
"**Zone Redundant Deployment**" is a Database that is deployed across multiple Availability Zones.
"**Primary**" means any Database that has active geo-replication relationship with a Database in other Azure regions. Primary can process read and write requests from the application.
"**Secondary**" means any Database that maintains asynchronous geo-replication relationship with a Primary in another Azure region and can be used as a failover target. Secondary can process read-only requests from applications.
"**Compliant Secondary**" means any Secondary that is created with the same configuration and in the same service tier as the Primary. If the Secondary is created in an elastic pool, it is considered Compliant if both Primary and Secondary are created in elastic pools with matching configurations and with density not exceeding 250 databases for a compliant configuration.

Uptime Calculation and Service Levels for Azure SQL Database Service
"**Deployment Minutes**" is the total number of minutes that a given Database has been operational in Microsoft Azure during an Applicable Period.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes for a given Microsoft Azure subscription during an Applicable Period.
Downtime: is the total accumulated Deployment Minutes across all Databases in a given Microsoft Azure subscription during which the Database is unavailable. A minute is considered unavailable for a given Database if all continuous attempts by Customer to establish a connection to the Database within the minute fail.
Uptime Percentage: for a given Database is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for a given Microsoft Azure subscription.
The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the General Purpose, Business Critical, Premium or Hyperscale tiers of the SQL Database Service configured for Zone Redundant Deployments:

Uptime Percentage	Service Credit
< 99.995%	10%
< 99%	25%
< 95%	100%

The following Service Levels and Service Credits are applicable to Customer's use of the Hyperscale, Business Critical, Premium or General Purpose, of the SQL Database Service not configured for Zone Redundant Deployments:

Uptime Percentage	Service Credit
< 99.99%	10%
< 99%	25%
< 95%	100%

The following Service Levels and Service Credits are applicable to Customer's use of the Basic or Standard tiers of the SQL Database Service:

Uptime Percentage	Service Credit
< 99.99%	10%
< 99%	25%
< 95%	100%

Recovery Point Objective (RPO)
"**Geo-Replication Link**" is a programmatic object representing a connection between a specific Primary and the Secondary.
"**Geo-Replication Lag**" is a time span from the point of transaction commit on the Primary and the acknowledgement by the Secondary that the transaction log update has been persisted.

"**Replication Lag Check**" is a programmatic method of obtaining the Geo-Replication Lag value for a specific Geo-Replication Link.
"**Recovery Point Objective (RPO)**" means a Geo-Replication Lag not to exceed 5 seconds.
"**N**" is the number of Replication Lag Check for a given Geo-Replication Link in a given hour.
"**S**" is the lag-sorted set of Replication Lag Check results in ascending order for a given Geo-Replication Link in a given hour.
"**Ordinal Rank**" is the 99th percentile using the nearest rank method represented by the following formula:

$$\frac{99}{100} x N$$

"**P99 Replication Lag**" is the value at the Ordinal Rank of S.
"**Deployment Hours**" is the total number of hours that a given Compliant Secondary has been operational for a given Microsoft Azure subscription during an Applicable Period.
"**Excessive Lag Hours**" is the total number of one-hour intervals during which Replication Lag Check resulted in a P99 Replication Lag greater than or equal to RPO for a given Microsoft Azure subscription during an Applicable Period. If the number of Replication Lag Checks in a given one-hour interval is zero, the Excessive Lag Hours for that interval is 0.
"**RPO Attainment Percentage**" for a given Database deployment in an Applicable Period is calculated using the following formula:

$$100\% - \frac{\text{Excessive Lag Hours}}{\text{Deployment Hours}} x 100$$

The following Service Levels and Service Credits are applicable to Customer's use of the active geo-replication feature with Business Critical tier of Azure SQL Database service with a Compliant Secondary:

Operation	RPO	RPO Attainment Percentage	Service Credit
Geo-Replication	5 seconds	< 100%	10% of total Applicable Period cost of Compliant Secondary

Recovery Time Objective (RTO)
"**Unplanned Failover**" is an action initiated by Customer when the Primary is offline to enable a Compliant Secondary as Primary.
"**Recovery Time**" is the time elapsed from the Unplanned Failover until the Secondary is acting as the Primary.
"**Recovery Time Objective (RTO)**" means a maximum allowed Recovery Time not to exceed 30 seconds.
"**Non-compliant Unplanned Failover**" is an Unplanned Failover that failed to complete within the RTO.
"**RTO Attainment Percentage**" for a given Database deployment, in an Applicable Period for a given subscription is represented by the following formula:

$$\frac{\text{Total Number of Unplanned Failovers - Total Number of Non-Compliant Unplanned Failo}}{\text{Total Number of Unplanned Failovers}}$$

The following Service Levels and Service Credits are applicable to Customer's use of the active geo-replication feature with Business Critical service tier of SQL Database service with a Compliant Secondary:

Operation	RTO	RTO Attainment Percentage	Service Credit
Unplanned Failover of Single Database	30 seconds	< 100%	100% of total Applicable Period cost of Compliant Secondary

Table of Contents / Definitions

Azure SQL Managed Instance

Additional Definitions
"**Instance**" means any Microsoft Azure SQL Managed Instance created in any of the Service tiers and deployed as a single instance.
"**Compliant Networking Configuration**" means full set of required configurations of the Microsoft Azure Virtual Network hosting Instance, including Microsoft Azure Network Security Group inbound security rules and mandatory Microsoft Azure User Defined Routes of Microsoft Azure Virtual Network Subnet hosting Instance, allowing uninterrupted flow of the management traffic and allowing data traffic to the dedicated gateway placed in the Microsoft Azure Virtual Network Subnet hosting Instance.
"**Availability Zone**" is a fault-isolated area within an Azure region, providing redundant power, cooling, and networking.

"**Zone Redundant Deployment**" is an Instance that is deployed across multiple Availability Zones.

Uptime Calculation and Service Levels for Azure SQL Managed Instance Service

"**Deployment Minutes**" is the total number of minutes that a given Instance has been operational in Microsoft Azure during an Applicable Period.

"**Maximum Available Minutes**" is the sum of all Deployment Minutes for a given Microsoft Azure subscription during an Applicable Period.

"**Downtime**" is the total accumulated Deployment Minutes across all Instances in a given Microsoft Azure subscription during which the Instance is unavailable. A minute is considered unavailable for a given Instance if all continuous attempts by Customer to establish a connection to the Instance within the minute fail.

"**Uptime Percentage**" for a given Instance is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for a given Microsoft Azure subscription. Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Business Critical tier of the SQL Managed Instance Service configured for Zone Redundant Deployments with Compliant Network Configuration:

Uptime Percentage	Service Credit
< 99.995%	10%
< 99%	25%
< 95%	100%

The following Service Levels and Service Credits are applicable to Customer's use of Business Critical tier of the SQL Managed Instance Service not configured for Zone Redundant Deployments with Compliant Network Configuration:

Uptime Percentage	Service Credit
< 99.99%	10%
< 99%	25%
< 95%	100%

The following Service Levels and Service Credits are applicable to Customer's use of the General Purpose tier of the SQL Managed Instance Service configured for Zone Redundant Deployments with Compliant Network Configuration:

Uptime Percentage	Service Credit
< 99.995%	10%
< 99%	25%
< 95%	100%

The following Service Levels and Service Credits are applicable to Customer's use of the General Purpose tier of the SQL Managed Instance Service not configured for Zone Redundant Deployments with Compliant Network Configuration:

Uptime Percentage	Service Credit
< 99.99%	10%
< 99%	25%
< 95%	100%

Table of Contents / Definitions

SQL Server Stretch Database

Additional Definitions:

"**Database**" means one instance of SQL Server Stretch Database.

"**Maximum Available Minutes**" is the total number of minutes that a given Database has been deployed in a given Microsoft Azure subscription during an Applicable Period.

Downtime: is the total accumulated minutes across all Databases deployed by Customer in a given Microsoft Azure subscription during which the Database is unavailable. A minute is considered unavailable for a given Database if all continuous attempts by Customer to establish a connection to the Database within the minute fail.

Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Static Web Apps

Additional Definitions
"**Deployment Minutes**" is the total number of minutes that a given App has been set to running in Microsoft Azure during an Applicable Period. Deployment Minutes is measured from when the App was created or Customer initiated an action that would result in running the App to the time Customer initiated an action that would result in stopping or deleting the App.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Apps deployed by Customer in a given Microsoft Azure subscription during an Applicable Period.
"**App**" is a web app deployed by Customer within the Static web apps.
Downtime: The total accumulated Deployment Minutes, across all Apps deployed by Customer in a given Microsoft Azure subscription, during which the App is unavailable. A minute is considered unavailable for a given App when there is no connectivity between the App and Microsoft's Internet gateway.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit

Uptime Percentage	Service Credit
< 99.95%	10%
< 99%	25%

Additional Terms: Service Credits are applicable only to fees attributable to your use of static web apps and not to fees attributable to other types of apps available.

Table of Contents / Definitions

Storage Accounts

Additional Definitions:
“**Archive Access Tier**” is a tier optimized for storing data that is rarely accessed, and that has flexible latency requirements, on the order of hours.
“**Average Error Rate**” for an Applicable Period is the sum of Error Rates for each hour in the Applicable Period divided by the total number of hours in the Applicable Period.
“**Blob Storage Account**” is a storage account specialized for storing data as blobs and provides the ability to specify an access tier indicating how frequently the data in that account is accessed.
“**Block Blob Storage Account**” is a storage account specialized for storing data as block or append blobs on solid-state drives.
“**Cold Access Tier**” is an attribute of a blob or account indicating it is rarely accessed and has a lower availability service level than blobs in Hot Access Tier.
“**Cool Access Tier**” is an attribute of a blob, file share, or account indicating it is infrequently accessed and has a lower availability service level than blobs in Hot Access Tier.
“**Hot Access Tier**” is an attribute of a blob, file share, or account indicating it is frequently accessed.
“**Excluded Transactions**” are storage transactions that do not count toward either Total Storage Transactions or Failed Storage Transactions. Excluded Transactions include pre-authentication failures; authentication failures; attempted transactions for storage accounts over their prescribed quotas; creation or deletion of containers, file shares, tables, or queues; clearing of queues; and copying blobs or files between storage accounts.
“**Error Rate**” is the total number of Failed Storage Transactions divided by the Total Storage Transactions during a set time interval (currently set at one hour). If the Total Storage Transactions in a given one-hour interval is zero, the error rate for that interval is 0%.

"**Failed Storage Transactions**" is the set of all storage transactions within Total Storage Transactions that are not completed within the Maximum Processing Time associated with their respective transaction type, as specified in the table below. Maximum Processing Time includes only the time spent processing a transaction request within the Storage Service and does not include any time spent transferring the request to or from the Storage Service.

Transaction Types	Maximum Processing Time
PutBlob and GetBlob (includes blocks and pages) Get Valid Page Blob Ranges	Two (2) seconds multiplied by the number of MBs transferred in the course of processing the request
PutFile and GetFile	Two (2) seconds multiplied by the number of MBs transferred in the course of processing the request
Copy Blob	Ninety (90) seconds (where the source and destination blobs are within the same storage account)
Copy File	Ninety (90) seconds (where the source and destination files are within the same storage account)
PutBlockList GetBlockList	Sixty (60) seconds
Table Query List Operations Find Operations	Ten (10) seconds (to complete processing or return a continuation)
Batch Table Operations	Thirty (30) seconds
All Single Entity Table Operations All other Blob, File, and Message Operations	Two (2) seconds

These figures represent maximum processing times. Actual and average times are expected to be much lower.

Failed Storage Transactions do not include:

1. Transaction requests that are throttled by the Storage Service due to a failure to obey appropriate back-off principles.
2. Transaction requests having timeouts set lower than the respective Maximum Processing Times specified above.
3. Read transactions requests to RA-GRS and RA-GZRS Accounts for which you did not attempt to execute the request against Secondary Region associated with the storage account if the request to the Primary Region was not successful.
4. Read transaction requests to RA-GRS and RA-GZRS Accounts that fail due to Geo-Replication Lag.

"**Geo Replication Lag**" for GRS, GZRS, RA-GRS, and RA-GZRS Accounts is the time it takes for data stored in the Primary Region of the storage account to replicate to the Secondary Region of the storage account. Because GRS, GZRS, RA-GRS, and RA-GZRS Accounts are replicated asynchronously to the Secondary Region, data written to the Primary Region of the storage account will not be immediately available in the Secondary Region. You can query the Geo Replication Lag for a storage account, but Microsoft does not provide any guarantees as to the length of any Geo Replication Lag under this SLA. "**Geo Redundant Storage (GRS) Account**" is a storage account for which data is replicated synchronously within a Primary Region and then replicated asynchronously to a Secondary Region. You cannot directly read data from or write data to the Secondary Region associated with GRS Accounts.
"**Locally Redundant Storage (LRS) Account**" is a storage account for which data is replicated synchronously only within a Primary Region.
"**Primary Region**" is a geographical region in which data within a storage account is located, as selected by you when creating the storage account. You may execute write requests only against data stored within the Primary Region associated with storage accounts.
"**Read Access Geo Redundant Storage (RA-GRS) Account**" is a storage account for which data is replicated synchronously within a Primary Region and then replicated asynchronously to a Secondary Region. You can directly read data from, but cannot write data to, the Secondary Region associated with RA-GRS Accounts.
"**Secondary Region**" is a geographical region in which data within a GRS or RA-GRS Account is replicated and stored, as assigned by Microsoft Azure based on the Primary Region associated with the storage account. You cannot specify the Secondary Region associated with storage accounts.
"**Total Storage Transactions**" is the set of all storage transactions, other than Excluded Transactions, attempted within a one-hour interval across all storage accounts in the Storage Service in a given subscription.
"**Transaction Optimized Access Tier**" is an attribute of an Azure file share indicating it is accessed frequently.
"**Zone Redundant Storage (ZRS) Account**" is a storage account for which data is replicated across multiple facilities. These facilities may be within the same geographical region or across two geographical regions.
"**Geo Zone Redundant Storage (GZRS) Account**" is a storage account for which data is replicated across multiple facilities. These facilities may be within the same geographical region or across two geographical regions. Data is also

replicated synchronously within a Primary Region and then replicated asynchronously to a Secondary Region. You cannot directly read data from or write data to the Secondary Region associated with GZRS Accounts.
"**Read Access Geo Zone Redundant Storage (RA-GZRS) Account**" is a storage account for which data is replicated across multiple facilities. These facilities may be within the same geographical region or across two geographical regions. Data is also replicated synchronously within a Primary Region and then replicated asynchronously to a Secondary Region. You can directly read data from, but cannot write data to, the Secondary Region associated with RA-GZRS Accounts.
Uptime Percentage: Uptime Percentage is calculated using the following formula:

$$100\% - Average\,Error\,Rate$$

Hot, and Transaction Optimized Access Tiers
Service Credit –LRS, ZRS, GRS, GZRS, RA-GRS, and RA-GZRS (write requests) for Hot, and Transaction Optimized Access Tiers:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Service Credit –RA-GRS and RA-GZRS (read requests) for Hot, and Transaction Optimized Access Tiers

Uptime Percentage	Service Credit
< 99.99%	10%
< 99%	25%

Service Credit – LRS, ZRS, GRS and GZRS (write requests) for Hot, and Transaction Optimized Access Tiers

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Cool, Cold, and Archive Tiers
Service Credit – LRS, ZRS, GRS, GZRS, RA-GRS, & RA-GZRS (write requests) for Cool, Cold, and Archive Access Tiers

Uptime Percentage	Service Credit
< 99%	10%
< 98%	25%

Service Credit – RA-GRS & RA-GZRS (read requests) Cool, Cold, and Archive Access Tiers

Uptime Percentage	Service Credit
< 99.9%	10%
< 98%	25%

Service Credit – LRS, ZRS, GRS, GZRS (read requests) for Cool, Cold, and Archive Access Tiers

Uptime Percentage	Service Credit
< 99.9%	10%
< 98%	25%

Service Exceptions: Cool, Cold and Archive SLA are applicable only to storage account types that support Cool, Cold and Archive tier.

Table of Contents / Definitions

StorSimple

Additional Definitions:
"**Backup**" is the process of backing up data stored on a registered StorSimple device to one or more associated cloud storage accounts within Microsoft Azure.
"**Cloud Tiering**" is the process of transferring data from a registered StorSimple device to one or more associated cloud storage accounts within Microsoft Azure.
"**Failure**" means the inability to fully complete a properly configured Backup, Tiering, or Restoring operation due to unavailability of the StorSimple Service.
"**Managed Item**" refers to a volume that has been configured to Backup to the cloud storage accounts using the StorSimple Service.

"**Restoring**" is the process of copying data to a registered StorSimple device from its associated cloud storage account(s).

Uptime Calculation and Service Levels for StorSimple Service

"**Deployment Minutes**" is the total number of minutes during which a Managed Item has been configured by Customer for Backup or Cloud Tiering to a StorSimple storage account in Microsoft Azure.

"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Managed Items for a given Microsoft Azure subscription during an Applicable Period.

Downtime: The total number of minutes within Maximum Available Minutes during which the StorSimple Service is unavailable for the Managed Item. The StorSimple Service is considered unavailable for a given Managed Item from the first Failure of a Backup, Cloud Tiering, or Restoring operation with respect to the Managed Item until the initiation of a successful Backup, Cloud Tiering, or Restoring operation of the Managed Item, provided that retries are continually attempted no less frequently than once every thirty minutes.

Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the StorSimple Service:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Uptime Calculation and Service Levels for StorSimple Data Manager

"**Total Requests**" is the set of all requests, other than Excluded Requests, to perform operations against StorSimple Data Manager service during an Applicable Period for a given Microsoft Azure subscription.

"**Excluded Requests**" is the set of requests that result in an HTTP 4xx status code.

"**Failed Requests**" is the set of all requests within Total Requests that either return an Error Code or fail to return a Success Code within 60 seconds.

Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Total Requests-Failed Requests}}{\text{Total Requests}} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure Stream Analytics

Uptime Calculation for Stream Analytics API Calls

Additional Definitions:

"**Total Transaction Attempts**" is the total number of authenticated REST API requests to manage a streaming job within the Stream Analytics Service by Customer during an Applicable Period for a given Microsoft Azure subscription.

"**Failed Transactions**" is the set of all requests within Total Transaction Attempts that return an Error Code or otherwise do not return a Success Code within five minutes from Microsoft's receipt of the request.

"**Uptime Percentage**" for API calls within the Stream Analytics Service is represented by the following formula:

$$Monthly\,Uptime\% = \frac{Total\,Transaction\,Attempts - Failed\,Transactions}{Total\,Transaction\,Attempts}$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Uptime Calculation for Stream Analytics Jobs

Additional Definitions:

"**Deployment Minutes**" is the total number of minutes that a given job has been deployed within the Stream Analytics Service during an Applicable Period.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all jobs deployed by Customer in a given Microsoft Azure subscription during an Applicable Period.
Downtime is the total accumulated Deployment Minutes, across all jobs deployed by Customer in a given Microsoft Azure subscription, during which the job is unavailable. A minute is considered unavailable for a deployed job if the job is neither processing data nor available to process data throughout the minute.
Uptime Percentage for jobs within the Stream Analytics Service is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure Synapse Analytics

Additional Definitions:
Synapse SQL
"**Database**" means any Synapse SQL database.
"**Maximum Available Minutes**" means the total number of minutes that a given Database has been deployed in Microsoft Azure during an Applicable Period in a given Microsoft Azure subscription.
"**Client Operations**" means the set of all documented operations supported by Azure Synapse Analytics.
"**Downtime**" means the total accumulated minutes during an Applicable Period for a given Microsoft Azure subscription during which a given Database is unavailable. A minute is considered unavailable for a given Database if more than 1% of all Client Operations completed during the minute return an Error Code.
"**Uptime Percentage**" for a given Database means a calculation of Maximum Available Minutes less Downtime, divided by Maximum Available Minutes in an Applicable Period in a given Azure subscription.
Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Data Integration in Azure Synapse
"**Data Integration Resources**" means integration runtimes (including Azure and self-hosted Integration Runtimes), triggers, pipelines, data sets, and linked services created within an Azure Synapse workspace.
"**Activity Run**" means the execution or attempted execution of an activity.
Uptime Calculation for Data Integration API Calls
"**Total Requests**" means the set of all requests, other than Excluded Requests, to perform operations against Data Integration Resources during an Applicable Period for a given Microsoft Azure subscription.
"**Excluded Requests**" means the set of requests that result in an HTTP 4xx status code, other than an HTTP 408 status code.
"**Failed Requests**" means the set of all requests within Total Requests that either return an Error Code or an HTTP 408 status code or otherwise fail to return a Success Code within two minutes.
"**Uptime Percentage**" for the API calls made to the Data Integration Resources is calculated as Total Requests less Failed Requests divided by Total Requests in an Applicable Period for a given Microsoft Azure subscription.
Uptime Percentage is represented by the following formula:

$$\frac{\text{Total Requests-Failed Requests}}{\text{Total Requests}} x100$$

The following Service Credits are applicable to Customer's use of Data Integration API calls within the Synapse workspace:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Apache Spark in Azure Synapse Calculation for Spark Sessions
"**Spark Session**" is the launching of a new session to execute a job, interactive or batch mode. Excluding session failures due to user error such as session config or exhausted resources.
The following Service Credits are applicable to Customer's use of Spark within the Synapse workspace:

Uptime Percentage	Service Credit
< 99%	10%
< 95%	25%

Table of Contents / Definitions

Azure Time Series Insights

Additional Definitions
"**Environment**" is a Time Series Insights environment.
Uptime Calculation and Service Levels for Time Series Insights data plane API
"**Time Series Insights data plane API**" is an event analytics query API for Time Series Insights.
"**Request**" is any documented request supported by the Time Series Insights data plane APIs.
"**Failed Request**" is a Request that returns an Error Code.
"**Error Rate**" is the total number of Failed Requests divided by the total number of Requests, during a given one-minute interval, for all Environments within a given Microsoft Azure Subscription. If the user has not made any Requests within this minute, the Error Rate for that interval is 0%.
"**Average Error Rate**" for an Applicable Period is the sum of Error Rates for each minute in Applicable Period divided by the total number of minutes in the Applicable Period.
"**Availability Percentage**" for the Time Series Insights data plane API is calculated by subtracting from 100% the Average Error Rate for a given Microsoft Azure Subscription in an Applicable Period. Availability Percentage is represented by the following formula:

100%-Average Error Rate

The following Service Levels and Service Credits are applicable to Customer's use of the Time Series Insights data plane API:

Availability Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Traffic Manager Service

Additional Definitions:
"**Deployment Minutes**" is the total number of minutes that a given Traffic Manager Profile has been deployed in Microsoft Azure during an Applicable Period.
"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Traffic Manager Profiles deployed by you in a given Microsoft Azure subscription during an Applicable Period.
"**Traffic Manager Profile**" or "**Profile**" refers to a deployment of the Traffic Manager Service created by you containing a domain name, endpoints, and other configuration settings, as represented in the Management Portal.
"**Valid DNS Response**" means a DNS response, received from at least one of the Traffic Manager Service name server clusters, to a DNS request for the domain name specified for a given Traffic Manager Profile.
Downtime: The total accumulated Deployment Minutes, across all Profiles deployed by you in a given Microsoft Azure subscription, during which the Profile is unavailable. A minute is considered unavailable for a given Profile if all continual DNS queries for the DNS name specified in the Profile that are made throughout the minute do not result in a Valid DNS Response within two seconds.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 100%	10%
< 99.99%	25%
< 99%	100%

Table of Contents / Definitions

Trusted Signing

Additional Definitions:
"**Total Transaction Attempts**" is the total number of authenticated API sign requests by Customer during an Applicable Period for a given Microsoft Azure subscription.
"**Failed Transactions**" is the set of all authenticated API sign requests within Total Transaction Attempts that return a HTTP 5xx status code or fail to return a Success Code within 90 seconds.
Uptime Percentage: is calculated as Total Transaction Attempts less Failed Transactions, divided by Total Transaction Attempts in an Applicable Period for a given Microsoft Azure subscription. Uptime Percentage is represented by the following formula:

$$\frac{\text{Total Transaction Attempts-Failed Transactions}}{\text{Total Transaction Attempts}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of Trusted Signing:
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Virtual Machines

Additional Definitions:
"**Availability Set**" refers to two or more Virtual Machines deployed across different Fault Domains to avoid a single point of failure.
"**Availability Zone**" is a fault-isolated area within an Azure region, providing redundant power, cooling, and networking.
"**Azure Dedicated Host**" provides physical servers that host one or more Azure virtual machines with the (default) setting of autoReplaceOnFailure required for any SLA.
"**Data Disk**" is a persistent virtual hard disk, attached to a Virtual Machine, used to store application data.
"**Dedicated Host Group**" is a collection of Azure Dedicated Hosts deployed within an Azure region across different Fault Domains to avoid a single point of failure.
"**Fault Domain**" is a collection of servers that share common resources such as power and network connectivity.
"**Operating System Disk**" is a persistent virtual hard disk, attached to a Virtual Machine, used to store the Virtual Machine's operating system.
"**Shared Disk**" is a Data Disk attached to multiple Virtual Machines simultaneously.
"**Single-Instance Virtual Machine**" is defined as any single Microsoft Azure Virtual Machine that either is not deployed in an Availability Set or has only one instance deployed in an Availability Set.
"**Virtual Machine**" refers to persistent instance types that can be deployed individually or as part of an Availability Set or using a Dedicated Host Group. A virtual machine can be deployed in a multi-tenant environment in Azure or in an isolated, single-tenant environment using Azure Dedicated Hosts.
"**Virtual Machine Connectivity**" is bi-directional network traffic between the Virtual Machine and other IP addresses using TCP or UDP network protocols in which the Virtual Machine is configured for allowed traffic. The IP addresses can be IP addresses in the same Cloud Service as the Virtual Machine, IP addresses within the same virtual network as the Virtual Machine or public, routable IP addresses.
Uptime Calculation and Service Levels for Virtual Machines in Availability Zones

"**Maximum Available Minutes**" is the total accumulated minutes during an Applicable Period that have two or more instances deployed across two or more Availability Zones in the same region. Maximum Available Minutes is measured from when at least two Virtual Machines across two Availability Zones in the same region have both been started resultant from action initiated by Customer to the time Customer has initiated an action that would result in stopping or deleting the Virtual Machines.

"**Downtime**" is the total accumulated minutes that are part of Maximum Available Minutes that have no Virtual Machine Connectivity in the region.
"**Uptime Percentage**" for Virtual Machines in Availability Zones is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for a given Microsoft Azure subscription. Uptime Percentage is represented by the following formula:

$$Monthly\,Uptime\% = \frac{(\text{Maximum Available Minutes-Downtime})}{\text{Maximum Available Minutes}} x100$$

Service Credit:
The following Service Levels and Service Credits are applicable to Customer's use of Virtual Machines deployed across two or more Availability Zones in the same region:

Uptime Percentage	Service Credit
< 99.99%	10%
< 99%	25%
< 95%	100%

Uptime Calculation and Service Levels for Virtual Machines in an Availability Set or in the same Dedicated Host Group

Maximum Available Minutes: The total accumulated minutes during an Applicable Period for all Internet facing Virtual Machines that have two or more instances deployed in the same Availability Set on in the same Dedicated Host Group. Maximum Available Minutes is measured from when at least two Virtual Machines in the same Availability Set, or same Dedicated Host Group, have both been started resultant from action initiated by you to the time you have initiated an action that would result in stopping or deleting the Virtual Machines.
Downtime: The total accumulated minutes that are part of Maximum Available Minutes that have no Virtual Machine Connectivity.
Uptime Percentage: for Virtual Machines is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for a given Microsoft Azure subscription. Uptime Percentage is represented by the following formula:

$$Monthly\,Uptime\% = \frac{(\text{Maximum Available Minutes-Downtime})}{\text{Maximum Available Minutes}} x100$$

Service Credit:
The following Service Levels and Service Credits are applicable to Customer's use of Virtual Machines in an Availability Set or same Dedicated Host Group. This SLA does not apply to Availability Sets leveraging Azure shared disks:

Uptime Percentage	Service Credit
< 99.95%	10%
< 99%	25%
< 95%	100%

Uptime Calculation and Service Levels for Single-Instance Virtual Machines and Virtual Machines using the same Shared Disks

"**Minutes in the Applicable Period**" is the total number of minutes in a given Applicable Period.
Downtime: is the total accumulated minutes that are part of Minutes in the Applicable Period that have no Virtual Machine Connectivity.
Uptime Percentage: is calculated by subtracting from 100% the percentage of Minutes in the Applicable Period in which any Single-Instance Virtual Machine had Downtime or in which all Virtual Machines using the same Shared Disk had Downtime.

$$Monthly\,Uptime\% = \frac{(\text{Minutes in the Applicable Period - Downtime})}{\text{Minutes in the Applicable Period}} x100$$

Service Credit:
The following Service Levels and Service Credits are applicable to Customer's use of Single-Instance Virtual Machines and Virtual Machines using the same Shared Disks by Disk type. For any Single-Instance Virtual Machine using multiple disk types and for all Virtual Machines using the same Shared Disks of multiple types*, the lowest SLA of all the disks on the Virtual Machine will apply:

*For example, given two Virtual Machines VM1 and VM2 using a Premium SSD Shared Disk and a Standard SSD Shared Disk, the uptime SLA for VM1 and VM2 will be the same as the SLA for a Single-Instance Virtual Machine using Standard SSD, as shown below.

Uptime Percentage (Premium SSD, Premium SSD v2, and Ultra Disk)**	Uptime Percentage (Standard SSD Managed Disk)	Uptime Percentage (Standard HDD Managed Disk)	Service Credit
< 99.9%	<99.5%	<95%	10%
< 99%	<95%	<92%	25%
< 95%	<90%	<90%	100%

**Premium SSD for all Operating System Disks, and Premium SSD, Premium SSD v2 or Ultra Disk for all Data Disks.

Table of Contents / Definitions

Azure Virtual Network Manager

Additional Definitions
"**Maximum Available Minutes**" is the total accumulated minutes during an Applicable Period during which a given Azure Virtual Network Manager has been deployed in a Microsoft Azure subscription.
"**Downtime**" is the total accumulated Maximum Available Minutes during which an Azure Virtual Network Manager is unavailable. A minute is considered unavailable if all attempts to connect to the Azure Virtual Network Manager within the minute are unsuccessful.
"**Uptime Percentage**" The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of each Azure Virtual Network Manager:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure Virtual WAN

Additional Definitions:
"**Maximum Available Minutes**" is the total accumulated minutes in an Applicable Period during which a given Azure Virtual WAN has been deployed in a Microsoft Azure subscription.
"**Downtime**" is the total accumulated Maximum Available Minutes during which an Azure Virtual WAN is unavailable. A given minute is considered unavailable if all attempts to connect to the Azure Virtual WAN throughout the minute are unsuccessful.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{Maximum\,Available Minutes - Downtime}{Maximum\,Available Minutes} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.95%	10%
< 99%	25%

Table of Contents / Definitions

Azure VMware Solution

Additional Requirements
Client is required to maintain a minimum configuration for all virtual machine storage including:

- For standard clusters: When the cluster has between 3 and 5 hosts, the number of failures to tolerate = 1; and when the cluster has between 6 and 16 hosts, the number of failures to tolerate = 2
- For stretched clusters: A minimum of 6 nodes are deployed in the cluster (3 in each availability zone), and you must have a minimum virtual machine storage policy of (i) primary level of failures to tolerate of "Dual-Site Mirroring" and (ii) a secondary level of failures to tolerate of 1 is used by the workload virtual machines.
- Storage capacity for the cluster retains slack space of 25% available (as described in the vSAN storage guide)
- Client has not performed any actions under the Elevated Privilege mode that prevent Microsoft from meeting the Availability Commitment.
- There is sufficient capacity on the cluster to support the Power-On of a virtual machine.
- Scheduled maintenance is excluded from the total available uptime calculations.

Additional Definitions
Uptime Calculation and Service Levels for Azure VMware Solution Workload Infrastructure
"**Maximum Available Minutes**" is the total accumulated minutes during an Applicable Period for all virtual machines within a VMware vSphere cluster, during which Azure VMware Solution has been deployed in a Microsoft Azure subscription.

"**Downtime**" is the total accumulated Maximum Available Minutes during an Applicable Period for a given VMware vSphere cluster on Azure, during which the Service is unavailable. A given minute is considered unavailable if any of the following are true:

- All Virtual Machines within a running cluster do not have any connectivity for four consecutive minutes.
- None of the Virtual machines can access storage for four consecutive minutes.
- None of the Virtual Machines can be started for four consecutive minutes.

"**Uptime Percentage**" The Uptime Percentage is calculated using the following formula:

$$\frac{Maximum\,Available Minutes - Downtime}{Maximum\,Available Minutes} x100$$

Service Credit

Uptime Percentage	Service Credit
< 99.9% for standard clusters	10%
<99.99% for stretched clusters	10%
< 99% any cluster	25%

Uptime Calculation and Service Levels for Azure VMware management tools

"**Maximum Available Minutes**" is the total accumulated minutes during an Applicable Period for a given VMware vSphere cluster, during which Azure VMware management tools have been deployed in a Microsoft Azure subscription.

"**Downtime**" is the total accumulated Maximum Available Minutes during an Applicable Period for a given VMware vSphere Cluster on Azure, during which the Management Services (vCenter Server and NSX Manager) are unavailable. A given minute is considered unavailable if any of the following are true:

- vCenter Server does not have any connectivity for four consecutive minutes.
- NSX Manager does not have any connectivity for four consecutive minutes.

"**Uptime Percentage**" The Uptime Percentage is calculated using the following formula:

$$\frac{Maximum\,Available Minutes - Downtime}{Maximum\,Available Minutes} x100$$

Service Credit

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Azure VMware Solution by CloudSimple

Additional Requirements

Client is required to maintain has a minimum configuration for all virtual machine storage as follows:

- When the cluster has between 3 and 5 hosts, the numbers of failure to tolerate = 1; and when the cluster has between 6 and 32 hosts, the number of failures to tolerate = 2
- Storage capacity for the cluster retains slack space of 25% available (as described in the VSAN storage guide). https://docs.vmware.com/en/VMware-vSphere/6.7/vsan-671-administration-guide.pdf
- There is sufficient capacity on the cluster to support the starting of a virtual machine, and Client has not performed any actions under the Escalated Privilege mode that prevent Vendor from meeting the Availability Commitment.
- Scheduled maintenance is excluded from the total available uptime calculations.

Additional Definitions

Uptime Calculation and Service Levels for Azure VMware Solutions Workload Infrastructure

"**Maximum Available Minutes**" is the total accumulated minutes during an Applicable Period for all virtual machines within a VMware cluster, during which Azure VMware Solutions has been deployed in a Microsoft Azure subscription.

"**Downtime**" is the total accumulated Maximum Available Minutes during an Applicable Period for a given VMware Cluster on Azure, during which the Service is unavailable. A given minute is considered unavailable if

- All Virtual Machines within a running cluster do not have any connectivity for four consecutive minutes.
- None of the Virtual machines can access storage for four consecutive minutes.
- None of the Virtual Machines can be started for four consecutive minutes.

"**Uptime Percentage**" The Uptime Percentage is calculated using the following formula:

$$\frac{Maximum\,Available Minutes - Downtime}{Maximum\,Available Minutes} x100$$

Service Credit

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	30%

Uptime Calculation and Service Levels for Azure VMware management tools

"**Maximum Available Minutes**" is the total accumulated minutes during an Applicable Period for a given VMware cluster, during which Azure VMware management tools have been deployed in a Microsoft Azure subscription.

"**Downtime**" is the total accumulated Maximum Available Minutes during an Applicable Period for a given VMware Cluster on Azure, during which the Management Services (vCenter Server and NSX Manager) is unavailable. A given minute is considered unavailable if

- vCenter server does not have any connectivity for four consecutive minutes.
- NSX Manager does not have any connectivity for four consecutive minutes.

"**Uptime Percentage**" The Uptime Percentage is calculated using the following formula:

$$\frac{Maximum\,Available Minutes - Downtime}{Maximum\,Available Minutes} x100$$

Service Credit

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	30%

Table of Contents / Definitions

Azure VNet NAT

Additional Definitions

"**Static Public IP address**" is an IP address that has been configured for a user's workload. A static IP address does not change.

"**Network address translation**" is the process of converting private IP addresses of a private network to a public IP address in order to allow multiple Azure compute resources (i.e., virtual machines) connect to the internet through a single public address.

"**Connectivity**" is bi-directional network traffic over supported IP transport protocols that can be sent and received from any IP address configured to allow traffic.

"**Outbound network traffic**" is traffic that is flowing from a private network to a public endpoint over the internet.

Uptime Calculation and Service Levels for Azure VNet NAT

"**Maximum Available Minutes**" is the total number of minutes that Azure VNet NAT (serving two or more Healthy Virtual Machines) has been deployed by Customer in a Microsoft Azure subscription during an Applicable Period.

"**Downtime**" is the total number of minutes within Maximum Available Minutes during which the given Azure VNet NAT is unavailable. A minute is considered unavailable if all Healthy Virtual Machines have no Connectivity through the VNet NAT endpoint. Downtime does not include minutes resulting from SNAT port exhaustions.

"**Uptime Percentage**" for Azure VNet NAT is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period multiplied by 100.

Uptime Percentage is represented by the following formula:

$$\frac{Maximum\,Available Minutes - Downtime}{Maximum\,Available Minutes} x100$$

The following Service Levels and Service Credit are applicable to Customer's use of the Azure NAT Gateway

Uptime Percentage	Service Credit
< 99.99%	10%
< 99.9%	25%

Table of Contents / Definitions

Virtual Network Gateway

Additional Definitions:

"**Maximum Available Minutes**" is the total accumulated minutes during an Applicable Period during which a given Virtual Network Gateway has been deployed in a Microsoft Azure subscription.

Downtime: Is the total accumulated Maximum Available Minutes during which a Virtual Network Gateway is unavailable. A minute is considered unavailable if all attempts to connect to the Virtual Network Gateway within a thirty-second window within the minute are unsuccessful.
Uptime Percentage: The Uptime Percentage for a given Virtual Network Gateway is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in an Applicable Period for the Virtual Network Gateway. The Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of each Virtual Network Gateway:
Basic Gateway for VPN or ExpressRoute Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Gateway for VPN and Gateway for ExpressRoute SKUs excluding Basic:

Uptime Percentage	Service Credit
< 99.95%	10%
< 99%	25%

Table of Contents / Definitions

Azure Web PubSub

Additional Definitions
"**Web PubSub service Endpoint**" is the host name from which the Web PubSub service is accessed by servers or clients to perform Web PubSub transactions.
"**Web PubSub Transactions**" is set of transaction requests sent from client to server or from server to client through a Web PubSub service Endpoint. These transaction requests include setting up the connection between server/client and Web PubSub service Endpoint or sending messages through the Web PubSub service Endpoint.
Uptime Calculation and Service Levels for Web PubSub service Instance
"**Maximum Available Minutes**" is the total number of minutes that the Web PubSub service has been deployed by the Customer in a given Microsoft Azure subscription during an Applicable Period.
"**Downtime**" is the total accumulated Maximum Available Minutes during an Applicable Period for the Web PubSub service during which the Web PubSub service is unavailable. A given minute is considered unavailable if all attempts to send Web PubSub transactions throughout the minute either return an Error Code or do not result in a Success Code within one minute.
"**Uptime Percentage**" for the Web PubSub service is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes.
Uptime Percentage is represented by the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Web PubSub service Standard tiers.

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Windows 10 IoT Core Services

Uptime Calculation and Service Levels for Windows 10 IoT Core Services
"**Maximum Available Minutes**" is the total number of minutes for a given Windows 10 IoT Core Services deployed by the Customer in a Microsoft Azure subscription during an Applicable Period.
"**Downtime**" is the total number of minutes within the Maximum Available Minutes during which Windows 10 IoT Core Services is unavailable. A minute is considered unavailable for a given Windows 10 IoT Core Services instance if all

continuous attempts to sign into Device Update Center or perform an operation within Device Update Center throughout the minute either return an Error Code or do not result in a Success Code within two minutes.
"**Uptime Percentage**" The Uptime Percentage is calculated using the following formula:

$$\frac{Maximum\,Available\,Minutes - Downtime}{Maximum\,Available\,Minutes} x100$$

The following Service Levels and Service Credits are applicable to Customer's use of Windows 10 IoT Core Services

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Other Online Services

Microsoft Defender for Identity

Additional Definitions:
"**Downtime**" is Any period of time when the admin is unable to access the Microsoft Defender for Identity portal.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Microsoft Defender for IoT

Additional Definitions:
"**Maximum Available Minutes**" represents the total number of minutes accumulated during an Applicable Period for the Microsoft Defender for IoT portal. Maximum Available Minutes are calculated from the time a Subscription is created following the successful completion of the onboarding process.
"**Downtime**" The total number of minutes that a customer is unable to access any part of a Microsoft Defender for IoT portal for which they have appropriate permissions, and have an active, valid license.
"Subscription" indicates the cloud environment specific to a particular customer for Microsoft Defender for IoT.
"Uptime Percentage" Downtime is measured in user minutes. That is, for an Applicable Period, Downtime will be the sum of the duration (in minutes) of each Incident during that Applicable Period multiplied by the number of users affected by that Incident. The Uptime Percentage is calculated using the following formula:

$$\frac{\text{Maximum Available Minutes-Downtime}}{\text{Maximum Available Minutes}} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Service Level Exceptions: This SLA does not apply to on-premises components such as sensors and security agents.

Table of Contents / Definitions

Bing Maps Enterprise Platform

Downtime: Any period of time when the Service is not available as measured in Microsoft's data centers, provided that you access the Service using the methods of access, authentication and tracking methods documented in the Bing Maps Platform SDKs.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{Total\,number\,of\,minutes \in an\,Applicable\,Period - Downtime}{Total\,number\,of\,minutes \in an\,Applicable\,Period} x100$$

where Downtime is measured as the total number of minutes during the Applicable Period when the aspects of the Service set forth above are unavailable.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Service Level Exceptions: This SLA does not apply to Bing Maps Enterprise Platform purchased through Open Value and Open Value Subscription volume licensing agreements.

Service Credits will not apply if: (i) you fail to implement any Services updates within the time specified in the Bing Maps Platform API's Terms of Use; and (ii) you do not provide Microsoft with at least ninety (90) days' advance notice of any known significant usage volume increase, with significant usage volume increase defined as 50% or more of the previous Applicable Period's usage.

Table of Contents / Definitions

Bing Maps Mobile Asset Management

Downtime: Any period of time when the Service is not available as measured in Microsoft's data centers, provided that you access the Service using the methods of access, authentication and tracking methods documented in the Bing Maps Platform SDKs.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{Total\,number\,of\,minutes \in an\,Applicable\,Period - Downtime}{Total\,number\,of\,minutes \in an\,Applicable\,Period} x100$$

where Downtime is measured as the total number of minutes during the Applicable Period when the aspects of the Service set forth above are unavailable.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Service Level Exceptions: This SLA does not apply to Bing Maps Enterprise Platform purchased through Open Value and Open Value Subscription volume licensing agreements.

Service Credits will not apply if: (i) you fail to implement any Services updates within the time specified in the Bing Maps Platform API's Terms of Use; and (ii) you do not provide Microsoft with at least ninety (90) days' advance notice of any known significant usage volume increase, with significant usage volume increase defined as 50% or more of the previous Applicable Period's usage.

Table of Contents / Definitions

Microsoft Cloud App Security

Downtime: Any period of time when the Customer's IT administrator or users authorized by Customer are unable to log on with proper credentials. Scheduled Downtime will not exceed 10 hours per calendar year.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{\text{User Minutes - Downtime}}{\text{User Minutes}} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Service Level Exceptions: This Service Level does not apply to any: (i) On-premises software licensed as part of the Service subscription, or (ii) Internet-based services (excluding Microsoft Cloud App Security) that provide updates via API (application programming interface) to any services licensed as part of the Service subscription.

Table of Contents / Definitions

Dragon Copilot

Additional Definitions:

"Ambient Speech" means the ability to record a health care encounter and receive a summarized draft clinical document.

"Ambient Speech Downtime" means the Downtime applicable to Ambient Speech. The Ambient Speech Downtime is the period of time in minutes during which Ambient Speech is materially degraded or unavailable due to a crash, malfunction, or other service disruption for which Microsoft is responsible. Ambient Speech will be considered materially degraded where the service fails to return a response within 15 minutes under normal network conditions.

"Front-End Speech" means the ability to dictate information into the targeted application and trigger specific actions with explicit voice commands.

"Front-End Speech Downtime" means the Downtime applicable to Front-End Speech. The Front-End Speech Downtime is the period of time in minutes during which Front-End Speech is materially degraded or unavailable due to a crash, malfunction, or other service disruption for which Microsoft is responsible. Front-End Speech will be considered materially degraded where the service fails to return a response within 2 minutes, under normal network conditions.

Uptime Percentage: The Uptime Percentage for Ambient Speech is calculated using the following formula:

$$\frac{User\,Minutes - Ambient\,Speech\,Downtime}{User\,Minutes} x100$$

where Ambient Speech Downtime is measured in user-minutes; that is, for each Applicable Period, Ambient Speech Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.

The Uptime Percentage for Front-End Speech is calculated using the following formula:

$$\frac{User\,Minutes - Front\,End\,Speech\,Downtime}{User\,Minutes} x100$$

where Front-End Speech Downtime is measured in user-minutes; that is, for each Applicable Period, Front-End Speech Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.

Service Credit for Ambient Speech:

Uptime Percentage	Service Credit
< 99%	5%
< 98%	10%

Service Credit for Front-End Speech:

Uptime Percentage	Service Credit
< 99.9%	5%
< 98.5%	10%

Additional Terms:

Only one Service Credit is permitted for Dragon Copilot for an Applicable Period. Where a Customer is otherwise eligible for both a Front-End Speech Service Credit and an Ambient Speech Service Credit, Customer may claim the Service Credit of the greatest amount.

Table of Contents / Definitions

Microsoft Power Automate

Downtime: Any period of time when users' flows have no connectivity to Microsoft's Internet gateway.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{Total\,number\,of\,minutes \in a\,month - Downtime}{Total\,number\,of\,minutes \in a\,month} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Microsoft Power Pages

This SLA applies to the end user website that is published using Power Pages if the following criteria are met: site must be in production mode, have site visibility set as public, and be associated with a production environment.
Additional Definitions:
"Total Website Requests" is the total number of website requests made by authorized (authenticated or anonymous) end users to a Power Pages website during an Applicable Period.
"Failed Website Requests" is the total number of requests within Total Website Requests for which Power Pages sends an error response due to a system error within Power Pages. User errors, issues due to inadequate licensing, or errors that are a result of customer's configuration or customization are not included in Failed Website Requests.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{Total\,Website\,Requests - Failed\,Website\,Requests}{Total\,Website\,Requests} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%

Table of Contents / Definitions

Microsoft Intune

Downtime: Any period of time when the Customer's IT administrator or users authorized by Customer are unable to log on with proper credentials. Scheduled Downtime will not exceed 10 hours per calendar year.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%

Uptime Percentage	Service Credit
< 95%	100%

Service Level Exceptions: This Service Level does not apply to any: (i) On-premises software licensed as part of the Service subscription, or (ii) Internet-based services (excluding Microsoft Intune Service) that provide updates to any on-premise software licensed as part of the Service subscription.

Table of Contents / Definitions

Microsoft Kaizala Pro

Downtime: Any period of time when end users are unable to read or post message in organizations groups for which they have appropriate permissions.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Microsoft Power Apps

Downtime: Any period of time when users are unable to read or write any portion of data in Microsoft Power Apps to which they have appropriate permissions.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{Total\,number\,of\,minutes \in a\,month - Downtime}{Total\,number\,of\,minutes \in a\,month} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Microsoft Copilot Studio

Additional Definitions:
"Total Message Requests" is the total number of requests made by an end user to Copilot Studio during an Applicable Period.
"Failed Message Requests" are the total number of requests within Total Message Requests that Copilot Studio is unable to send a response message to due to a system error within Copilot Studio.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{Total\,Message\,Requests - Failed\,Message\,Requests}{Total\,Message\,Requests} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%

Table of Contents / Definitions

Microsoft Sustainability Manager

Downtime: Any period of time when end users are unable to log into their environment. Downtime does not include Scheduled Downtime, the unavailability of Service add-on features, or the inability to access the Service due to your modifications of the Service.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.5%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Minecraft: Education Edition

Downtime: Any period of time when users are unable to access Minecraft: Education Edition.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{Total\,number\,of\,minutes \in a\,month - Downtime}{Total\,number\,of\,minutes \in a\,month} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Power BI Embedded

Deployment Minutes: The total number of minutes for which a given embedded capacity has been active during an Applicable Period.
Maximum Available Minutes: The sum of all Deployment Minutes for a specific embedded capacity provisioned by a customer in a given Microsoft Azure subscription during an Applicable Period.
Downtime Minutes: The total accumulated Deployment Minutes during which an embedded capacity is unable to be utilized in all applicable Power BI features listed below:
- **View:** View Power BI Dashboards, Reports, and Apps in the service.
- **Dataset Refresh:** Schedule or manually trigger refresh operation and expect those operations to complete within expected timeframes considering all conditions that might impact refresh speeds (e.g., size of dataset).
- **Access Power BI Portal:** Access and use the Power BI Portal within expected timeframes considering network conditions and limitations local to the customer environment or external to Microsoft.

Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{Maximum\,Available\,Minutes - Downtime\,Minutes}{Maximum\,Available\,Minutes} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Power BI Premium

Capacity: Means a named capacity provisioned by an admin through the Power BI Premium capacity admin portal. A Capacity is a grouping of one or more nodes.
Maximum Available Minutes: The sum of all minutes that a given Capacity has been instantiated during an Applicable Period for a given tenant.
Downtime Minutes: The total accumulated minutes in an Applicable Period for a given Capacity, after its creation, or before it is deprovisioned when the Capacity is unable to be utilized in all applicable Power BI features listed below:
View: View Power BI Dashboards, Reports, and Apps in the service.
Dataset Refresh: Schedule or manually trigger refresh operation and expect those operations to complete within expected timeframes considering all conditions that might impact refresh speeds (e.g., size of dataset).
Access Power BI Portal: Access and use the Power BI Portal within expected timeframes considering network conditions and limitations local to the customer environment or external to Microsoft.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{Maximum\,Available\,Minutes - Downtime\,Minutes}{Maximum\,Available\,Minutes} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Table of Contents / Definitions

Power BI Pro

Downtime Minutes: The total accumulated minutes in an Applicable Period during which all Power BI features listed below are unavailable:
View: View Power BI Dashboards, Reports, and Apps in the service.
Dataset Refresh: Schedule or manually trigger refresh operation and expect those operations to complete within expected timeframes considering all conditions that might impact refresh speeds (e.g., size of dataset).
Access Power BI Portal: Access and use the Power BI Portal within expected timeframes considering network conditions and limitations local to the customer environment or external to Microsoft.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{Total\,number\,of\,minutes \in a\,month - Downtime\,Minutes}{Total\,number\,of\,minutes \in a\,month} x100$$

Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Azure AI Translator

Downtime: Any period of time when users are not able to perform translations.

Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{Total\,number\,of\,minutes \in an\,Applicable\,Period - Downtime}{Total\,number\,of\,minutes \in an\,Applicable\,Period} x100$$

where Downtime is measured as the total number of minutes during the Applicable Period when the aspects of the Service set forth above are unavailable.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Table of Contents / Definitions

Microsoft Defender for Endpoint

Additional Definitions:
"**Maximum Available Minutes**" is the total accumulated minutes during an Applicable Period for Microsoft Defender for Endpoint portal. Maximum Available Minutes is measured from when the Tenant has been created resultant from successful completion of the on-boarding process.
"**Tenant**" represents Microsoft Defender for Endpoint customer specific cloud environment.
Downtime: The total accumulated minutes that are part of Maximum Available Minutes in which the Customer unable to access any portion of a Microsoft Defender for Endpoint portal site collections for which they have appropriate permissions and customer has a valid, active, license.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{Maximum\,Available\,Minutes - Downtime}{Maximum\,Available\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	10%
< 99%	25%

Service Level Exceptions: This SLA does not apply to any trial/preview version Tenants.

Table of Contents / Definitions

Universal Print

Downtime: Any period of time when unavailability of the Universal Print Service results in the inability for users to discover printers or submit print jobs, or the inability for administrators to register or configure printers, manage access control, or monitor Universal Print status and usage.
Uptime Percentage: The Uptime Percentage is calculated using the following formula:

$$\frac{User\,Minutes - Downtime}{User\,Minutes} x100$$

where Downtime is measured in user-minutes; that is, for each Applicable Period, Downtime is the sum of the length (in minutes) of each Incident that occurs during that Applicable Period multiplied by the number of users impacted by that Incident.
Service Credit:

Uptime Percentage	Service Credit
< 99.9%	25%
< 99%	50%
< 95%	100%

Service Level Exceptions: This SLA does not apply to any trial/preview version Tenants.

Table of Contents / Definitions

Windows 365

Cloud PC: the specific instance of Windows 365 licensed to a user.
Downtime: measured in minutes, the period in which all connection attempts by a specific user to a specific Cloud PC were unsuccessful, excluding any of the following types of failures:

1. Failures resulting from the Cloud PC being in an inoperable state unrelated to the underlying Azure infrastructure (e.g., damaged or corrupt operating system, operating system configuration, or misconfiguration); and
2. Failure resulting from an application or other software installed on the Cloud PC.

Individual Downtime: means Downtime for a given user for each Applicable Period.

Individual Minutes: means the User Minutes for a given user for each Applicable Period.

Individual Uptime Percentage: Individual Uptime Percentage is calculated as:

$$\frac{Individual\,Minutes - Individual\,Downtime}{Individual\,Minutes} x100$$

Per-User Credit: For an Applicable Period in which the Regional Uptime Percentage is less than 99.9%, a Per-User Credit shall be calculated as a percentage of the per user portion of the Applicable Service Fees for each user for whom the Individual Uptime Percentage was less than 99.9% according to the following table (provided, however, that any Individual Uptime Percentage that is lower than the Regional Uptime Percentage shall be deemed to be equal to the Regional Uptime Percentage):

Individual Uptime Percentage	Per User Credit
< 99.9%	10%
< 99%	25%
< 95%	100%

Region: means the regions detailed at: https://aka.ms/DSLARegionLink.

Regional Downtime: means the sum of all your Downtime in a Region for each Applicable Period.

Regional Minutes: means the User Minutes in a Region for each Applicable Period.

Regional Uptime Percentage: is calculated using the following formula:

$$\frac{Regional\,Minutes - Regional\,Downtime}{Regional\,Minutes} x100$$

Service Credit: for Windows 365, Service Credits are not a percentage of the Applicable Service Fee, but shall be the sum of all Per-User Credits.

Table of Contents / Definitions

Appendix A – Service Level Commitment for Virus Detection and Blocking, Spam Effectiveness, or False Positive

With respect to Exchange Online and EOP licensed as a standalone Service or via ECAL suite, or Exchange Enterprise CAL with Services, you may be eligible for Service Credits if we do not meet the Service Level described below for: (1) Virus Detection and Blocking, (2) Spam Effectiveness, or (3) False Positive. If any one of these individual Service Levels is not met, you may submit a claim for a Service Credit. If one Incident causes us to fail more than one SLA metric for Exchange Online or EOP, you may only make one Service Credit claim for that incident per Service.

1. **Virus Detection and Blocking Service Level**
   a. "Virus Detection and Blocking" is defined as the detection and blocking of Viruses by the filters to prevent infection. "Viruses" is broadly defined as known malware, which includes viruses, worms, and Trojan horses.
   b. A Virus is considered known when widely used commercial virus scanning engines can detect the virus and the detection capability is available throughout the EOP network.
   c. Must result from a non-purposeful infection.
   d. The Virus must have been scanned by the EOP virus filter.
   e. If EOP delivers an email that is infected with a known virus to you, EOP will notify you and work with you to identify and remove it. If this results in the prevention of an infection, you won't be eligible for a Service Credit under the Virus Detection and Blocking Service Level.
   f. The Virus Detection and Blocking Service Level shall not apply to:
      i. Forms of email abuse not classified as malware, such as spam, phishing and other scams, adware, and forms of spyware, which due to its targeted nature or limited use is not known to the anti-virus community and thus not tracked by anti-virus products as a virus.
      ii. Corrupt, defective, truncated, or inactive viruses contained in NDRs, notifications, or bounced emails.
   g. The Service Credit available for the Virus Detection and Blocking Service is: 25% Service Credit of Applicable Service Fee if an infection occurs in an Applicable Period, with a maximum of one claim allowed per Applicable Period.

2. **Spam Effectiveness Service Level**
   a. "Spam Effectiveness" is defined as the percentage of inbound spam detected by the filtering system, measured on a daily basis.
   b. Spam effectiveness estimates exclude false negatives to invalid mailboxes.
   c. The spam message must be processed by our service and not be corrupt, malformed, or truncated.
   d. The Spam Effectiveness Service Level does not apply to email containing a majority of non-English content.
   e. You acknowledge that classification of spam is subjective and accept that we will make a good faith estimation of the spam capture rate based on evidence timely supplied by you.
   f. The Service Credit available for the Spam Effectiveness Service is:

% of Applicable Period that Spam Effectiveness is below 99%	Service Credit
>25%	25%
> 50%	50%
100%	100%

3. **False Positive Service Level**
   a. "False Positive" is defined as the ratio of legitimate business email incorrectly identified as spam by the filtering system to all email processed by the service in an Applicable Period.
   b. Complete, original messages, including all headers, must be reported to the abuse team.
   c. Applies to email sent to valid mailboxes only.
   d. You acknowledge that classification of false positives is subjective and understand that we will make a good faith estimation of the false positive ratio based on evidence timely supplied by you.
   e. This False Positive Service Level shall not apply to:
      i. bulk, personal, or pornographic email
      ii. email containing a majority of non-English content
      iii. email blocked by a policy rule, reputation filtering, or SMTP connection filtering
      iv. email delivered to the junk folder
   f. The Service Credit available for the False Positive Service is:

False Positive Ratio in an Applicable Period	Service Credit
> 1:250,000	25%
> 1:10,000	50%
> 1:100	100%

Table of Contents / Definitions